Persistent holes in the Universe

Pranav, Pratyush

university of groningen

Persistent Holes in the Universe

A hierarchical topology of the cosmic mass distribution

PhD Thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 18 December 2015 at 14.30 hours

by

Pratyush Pranav

born on 18 February 1985
in Darbhanga, India

Supervisors	:	Prof. dr. R. van de Weygaert Prof. dr. G. Vegter Prof. dr. B. J. T. Jones
Assessment committee	:	Prof. dr. P. Barthel Prof. dr. V. Sahni Prof. dr. L. Wasserman

Then was not non-existence nor existence
There was no realm of air, no sky beyond it
What covered in, and where? And what gave shelter?
Death was not then, nor was there aught immortal
No sign was there, the day's and night's divider

That One Thing, breathless, breathed by its own nature
Darkness there was at first
Concealed in darkness this all was indiscriminated chaos
All that existed then was void and formless
By the great power of Warmth was born that Unit

There were begetters, there were mighty forces
Free action here and energy up yonder
Who verily knows and who can here declare it
Whence it was born
Whence comes this creation?

– *The Rig Veda, 10.129.1–7*

Cover Illustration: An artistic impression of the filamentary network that pervades the large scale Universe. Superposed is the nearby Universe depicting the stars in the night sky.

Concept : Pratyush Pranav & Maryse Koenders
Design : Maryse Koenders

Printed by: ViBa Press Pvt. Ltd., New Delhi – 20

ISBN 978-90-367-8453-5
ISBN 978-90-367-8452-8 (electronic version)

Contents

1

Introduction

Cosmology has come of age. During the last few decades, it has transformed from an endeavour, where it was all about the search for 'the two numbers' – the *Hubble parameter* and the *deceleration parameter*, to a stage where these two numbers are eventually expected to be constrained as a byproduct of the surveys aimed at a more detailed understanding of the origin, structure and dynamics of the Cosmos (Jones 1992; Kim et al. 1995; Hoeflich & Khokhlov 1996; Neben & Turner 2013). Due to recent advances in observational cosmology, humanity is at a stage where we have a broad understanding of how the Universe came into being and what its composition is. According to the standard paradigm, the Universe started in the state of a hot *big bang*. The latest measurements of the Cosmic microwave Background radiation, as performed by the PLANCK satellite, establish that only 4.9% of the total mass-energy content of the Universe is composed of baryonic matter, of which the galaxies, stars and us humans are made. The rest of 26.8% and 68.3% is respectively composed of *dark matter* and *dark energy* (Planck Collaboration et al. 2015).

The recent advances, while establishing the broader picture of the Universe on the one hand, have given way to a whole lot of new questions, while simultaneously pointing to the discrepancies in our current understanding of the details of the physical contents and processes that govern the dynamics and evolution of the Universe. Prominent is the question about the nature of dark matter and dark energy, which remains a complete mystery to us. With growing observational data, discrepancies in our theories of the evolution and structural organization of the Universe on the large scale have also come to the forefront. Similarly, our understanding of the formation and evolution of galaxies remain nebulous still.

The effort to describe and characterize the formation and evolution of structures in the Universe has been ongoing for decades now. The tools used for this purpose

have drawn heavily from concepts in statistics, geometry and topology. In the next few sections, I will present an overview of the current standard paradigm of cosmology, as well as a brief but non-exhaustive account of structure formation and evolution in the Universe. I will also present a brief account of the various tools and formalisms used to describe the properties of the cosmic matter distribution. This will establish a background for the new topological formalisms that will be described in this thesis. Based on this, we will present a topological characterization of heuristic as well as more realistic models of mass distribution in the Universe. This is with a view to understand and demonstrate what such a new topological characterization has to offer over the more traditional topological measures that have been used to describe the cosmic mass distribution in the past.

1.1 General Cosmology: an overview

As a result of continuous advances in theory and observations in tandem since almost the beginning of this century, the *Hot Big Bang Model* has established itself as the standard model of the origin and evolution of the Universe. According to this paradigm, we live in an expanding Universe that originated in a dense superhot state of singularity almost 13.8 Gyrs ago (Fixsen et al. 1994; Komatsu et al. 2009; Planck Collaboration et al. 2015).

The source of the current standard paradigm of cosmology, the idea that the Universe started in a hot Big Bang, may be attributed indirectly to Einstein who revolutionized our way of thinking about gravitation with his formulation of General Relativity (Einstein 1916). Culminating in the famous Einstein's field equations, General relativity interpreted the interaction of gravitation as a result of spacetime getting curved by the matter-energy content present in it. The Einstein field equations may be written as

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.1}$$

Each of the terms in the field equation is associated with either the curvature of the spacetime, or its matter energy content. In the above equation, $R_{\mu\nu}$ is the *Ricci curvature tensor*, $g_{\mu\nu}$ is the *metric tensor*, R is the *scalar curvature*, Λ is the *cosmological constant*, G is the *gravitational constant*, c is the speed of light and $T_{\mu\nu}$ is the *stress-energy tensor*. The left-hand side of the equation quantifies the curvature of spacetime, and the right-hand side quantifies its matter-energy content.

1.1.1 The cosmological principle and the Robertson-Walker Metric

The Einstein field equations are metric equations, meaning that the solution to them are metrics defined on a manifold. A *metric* captures the notion of distance on a given manifold. When written in the expanded form, the Einstein field equations are a set of ten coupled non-linear partial differential equations. Due to non-linearity, as well as coupling, it is not possible to find generic solutions to them. However, under assumptions of symmetry, homogeneity and isotropy, the set of equations

admit exact metric solution.

To build an understanding of the dynamics of the Universe, cosmologists start with the assumption that there are no preferred directions or preferred locations in the Universe at large scales. This assumption is encapsulated in the *cosmological principle*, which states that the Universe is isotropic and homogeneous Additionally, if we assume that the geometry of the spacetime is uniformly curved, there are only three possibilities for such a spacetime – it can be positively curved or have a *spherical* geometry; it can be negatively curved or have a *hyperbolic* geometry; or it can have zero curvature or a *flat* or the familiar *Euclidean* geometry. The metric for a homogeneous, isotropic and a uniformly curved spacetime is given by the *Robertson-Walker metric*. In spherical coordinates (r, θ, ϕ), this is

$$ds^2 = -c^2dt^2 + a(t)^2\left[dr^2 + S_k(r)^2 d\Omega^2\right], \tag{1.2}$$

where,

$$d\Omega^2 = d\theta^2 + sin^2\theta, d\phi^2, \tag{1.3}$$

and

$$S_k(r) = \begin{cases} R\sin(r/R), & k = +1, \text{closed universe} \\ r, & = 0, \text{ flat universe} \\ R\sinh(r/R), & k = -1, \text{ open universe} \end{cases}$$

The quantity R denotes the radius of curvature and the quantity $a(t)$ is the *expansion factor*. The latter relates the physical coordinate to the comoving coordinate via the relation r $= a(t)$R. By convention, $a(t_0) = 1$, where t_0 is the present time.

1.1.2 The FRWL equations

General relativity laid the foundation for Georges Lemaître to coin the idea of the Big Bang Universe – a universe that is not static, but has a finite age. Friedmann and Lemaître independently discovered the expanding Universe as a solution to the field equations under the assumption that the universe was isotropic and homogeneous (Friedmann 1922; Lemaître 1927). Known famously as the *Friedmann-Robertson-Walker-Lemaître* equations, they are given by

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3} \tag{1.4}$$

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda}{3}. \tag{1.5}$$

From the above equations, it is also possible to derive the *energy equation*

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right) = 0. \tag{1.6}$$

The energy equation is a statement of conservation of energy. It implies an adiabatically expanding Universe, which is also homogeneous and isotropic.
In the FRWL equations, G is the gravitational constant, ρ is an indicator of the energy density, p is the pressure, and Λ is the cosmological constant. It is usually the custom to express the FRWL equations in terms of the Hubble parameter H, and the density parameter Ω, such that

$$H^2(z) = H_0^2\left(\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda\right), \tag{1.7}$$

where,

$$a = 1/(1+z)H = \dot{a}/a. \tag{1.8}$$

In the above equation, z is the redshift corresponding to the scale factor a. At the present epoch, when matter is the dominating component, the density parameters are given by

$$\Omega_m = \frac{\rho_m}{\rho_c},\ \Omega_\Lambda = \frac{\Lambda}{3H_0^2},\ \Omega_k = -\frac{kc^2}{H_0^2} \tag{1.9}$$

where H_0 is the current value of the Hubble parameter The density of the various components are usually specified in the units of the critical density

$$\rho_c = 3H_0^2/8\pi G. \tag{1.10}$$

The critical density is the energy density for which the curvature of the Universe is flat.

Table 1.1 presents the latest values of the cosmological parameters as deduced from the measurement of the temperature anisotropies in the CMB by the PLANCK satellite (Planck Collaboration et al. 2015).

1.1.3 Evolution of the FRWL Universe

The FRWL equations fully describe the evolution of the Universe in terms of its constituents. The evolution and the eventual fate of the FRWL universe is fully determined by the

- Energy content of the universe, represented through the density and pressure terms.
- Geometry of the universe, parametrized through the curvature term.

Symbol	Value	Description
t_0	13.799±0.021 Gyr	Age of the Universe
H_0	67.74±0.46	Hubble parameter
$\Omega_b h^2$	0.02230±0.00014	Physical Baryon density
$\Omega_{DM} h^2$	0.1188± 0.0010	Physical dark Matter density
Ω_Λ	0.6911±0.0062	Dark energy density
σ_8	0.8159±0.0086	rms of density fluctuations at $8h^{-1}$Mpc
n_s	0.9667±0.0040	Scalar spectral index
τ	0.066±0.012	Reionization optical depth

Table 1.1 The latest values of the cosmological parameters as inferred from the measurement of the temperature anisotropies in the CMB by PLANCK satellite. Values taken from Planck Collaboration et al. (2015).

– The cosmological constant, denoted by Λ. The most popular interpretation of cosmological constant is as a form of energy which has negative pressure and contributes to an accelerated expansion of the Universe.

The energy content of the Universe has contributions from radiation, matter and dark energy. Matter consists of baryonic matter and dark matter. The different components have different contribution to the energy budget of the Universe. This is encapsulated in the *equation of state*, given by

$$p(\rho) = w\rho c^2, \tag{1.11}$$

where,

$$w = \begin{cases} 1/3 & \text{radiation} \\ 0 & \text{matter} \\ -1/3 < w < -1 & \text{dark energy.} \end{cases}$$

The case when $w = -1$ is associated with the cosmological constant. The negative pressure results in a repulsive force responsible for the accelerated expansion of the Universe. In fact, in order to get a term which causes an accelerated expansion, it is enough to have a scalar field which satisfies

$$p < -\frac{\rho c^2}{3}. \tag{1.12}$$

One option for such an (evolving) field is *quintessence* (Peebles & Ratra 1988), and leads to the notion of *dark energy*. One should note, however, that though the cosmological constant and the dark energy have similar behaviour in terms of the equation of state, their sources are different. The cosmological constant is associated with the curvature term on the left hand side of the Einstein field equations, and has its source in the curvature of the spacetime. Dark energy is associated with the energy term in the right hand side of the field equations.

At different cosmological epochs, different components were dominant. In the very early inflationary era, when the Universe is in a pseudo-vacuum state, and

expands exponentially, $w = -1$. This negative repulsive pressure is responsible for the exponential expansion of the Universe. The case $w = \frac{1}{3}$, is the era when radiation was the dominating component. At the current epoch, when matter is the dominant component, $w = 0$. However, it has been discovered recently that since approximately the last 7 Gyrs, the Universe is in a state of accelerated expansion (Riess et al. 1998; Perlmutter et al. 1999). This suggests that dark energy has started to dominate the dynamics of the Universe again about 7 billion years ago.

The evolution of the scale factor $a(t)$, is influenced differently by the different components of the Universe. Though all the components of the Universe contribute to the evolution of the scale factor, it turns out that over vast periods of time only one of the components is dominant. It is then instructive to study the time evolution of $a(t)$ by only considering the contribution from the dominating component and ignoring the contribution from the rest. It can be shown that during the different cosmological epochs, when different components are dominating, the time evolution of $a(t)$ is approximately given by

$$a(t) \propto \begin{cases} (t/t_0)^{1/2}, & \text{when radiation is dominant} \\ (t/t_0)^{2/3}, & \text{when matter is dominant} \\ e^{H_0(t-t_0)}, & \text{when dark energy is dominant} \end{cases}$$

Having assessed the dynamics and evolution of the model universe, we now present an account of the structure of the matter distribution in the Universe.

1.2 Structure formation in the Universe

Observations show that on the scales larger than a few hundred megaparsecs, the Universe appears homogeneous. However, at smaller scales, it shows very distinct features and structures. On the large scales, the structures manifest as a web-like network commonly known as the *cosmic web* (Bond et al. 1996). On yet progressively smaller scales, one finds the galaxies, stars and planets. The existence of these pronounced structures reveals that on these scales the Universe is no longer homogeneous and isotropic. These structures are the result of a growth of quantum fluctuations planted in the primordial density field in the very early Universe. The theoretical framework that describes this is that of gravitational instability.

1.2.1 Gravitational instability and linear growth of structures

The generally accepted theoretical framework for the formation of structure is that of gravitational instability (Peebles 1980; van de Weygaert & Bond 2008). This framework assumes that the early universe was almost perfectly smooth, with the exception of tiny density deviations with respect to the global cosmic background density, and the accompanying tiny velocity perturbations from the general Hubble expansion. The origin of these fluctuations is of a quantum nature. The most plausible reasoning is that they are the quantum fluctuations of the vacuum state, in which the Universe was, during the inflationary era.

The tiny local deviations start to grow under the influence of the accompanying gravity perturbations. In a perfectly homogeneous Universe the gravitational force is the same everywhere. The presence of density perturbations will induce local differences in gravity. In the vicinity of a region with a higher density than the universal background density, the surplus of matter will exert an attractive gravitational force larger than the average value. Similarly, near low density regions a deficit in matter will lead to a weaker force. This locally varying difference in the gravitational force induces a locally varying deceleration in the cosmic expansion. As it is the tendency of matter to move towards a region of higher density, the mass of the overdense regions increases with time.

If and when the region becomes sufficiently overdense, its expansion may even come to a halt. The region decouples completely from the Hubble expansion, turns around, and starts to contract. If or as long as pressure forces are not sufficient to counteract the infall, the overdensity will grow without bound, and assemble more and more matter by accretion from its surroundings. Ultimately this will turn into a fully collapsed, gravitationally bound object, which seeks to attain virial equilibrium. Once it has done so, a genuine identifiable cosmic object has formed (van de Weygaert & Bond 2008).

The opposite tendency occurs in the case of primordial density depressions. Because they contain less matter than on average, the deceleration of the matter in and around such an underdense region is less than that of the global Hubble expansion. Matter will therefore tend to get displaced somewhat further, with the net result of matter streaming out of the interior of the underdensities resulting in their expansion with respect to the global Universe. As the process continues and becomes more pronounced, the gravitational instability process results in the gradual emergence of a void in the matter distribution.

1.2.2 Linear structure growth

In the comoving coordinates, the time-dependent density contrast at any location in the Universe can be written as

$$\delta(\mathbf{x},t) \equiv \frac{\rho(\mathbf{x},t)-\rho_u(t)}{\rho_u(t)}, \tag{1.13}$$

where $\rho_u(t)$ is the universal background density at time t, and $\rho(\mathbf{x},t)$ is the density at comoving position $\mathbf{x}$ at time t. The story of structure growth in the Universe is the story of the evolution of over- and under-densities $\delta(\mathbf{x},t)$ across the cosmic timeline.

The location of an object in the Universe is specified by its physical coordinates $\mathbf{r}$. In an ideal FRW Universe, only the universal expansion is responsible for changing the physical coordinate of this object. As such, the physical coordinate can be expressed in terms of the comoving coordinates $\mathbf{x}$ and the expansion factor $a(t)$, such that

$$\mathbf{r}(t) = a(t)\mathbf{x}. \tag{1.14}$$

In the pure FRW Universe, where the density at any location is the same as the universal background density, **x** does not change with time. However, in the context of structure formation due to gravity perturbations, the comoving coordinate becomes a time dependent quantity $\mathbf{x}(t)$,

$$\mathbf{x}(t) = \frac{\mathbf{r}(t)}{a(t)}. \tag{1.15}$$

The evolution of density perturbations in an expanding Universe is governed by a set of three equations – the continuity equation, the Euler equation and the Poisson equation. The continuity equation is a statement of conservation of mass, while the Euler equation is a statement of conservation of momentum, and is basically the equation of motion. The Poisson equation relates the gravitational field to its source, the mass distribution in the Universe. They are given respectively by

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \frac{1}{a}\vec{\nabla}_x \cdot (1+\delta)\vec{v} &= 0, \\ \frac{\partial \vec{v}}{\partial t} + \frac{1}{a}\left(\vec{v}\cdot\vec{\nabla}_x\right)\vec{v} + \frac{\dot{a}}{a}\vec{v} &= -\frac{1}{a}\vec{\nabla}_x\phi \\ \nabla_x^2\phi &= 4\pi G a^2 \rho_u \delta. \end{aligned} \tag{1.16}$$

In the above equations, a is the expansion factor, $\vec{v}$ is the peculiar velocity, and ϕ is the gravitational potential, all expressed in comoving coordinates.

There are three distinct structural regimes in the Universe that can be demarcated on based on the value of δ. These are the linear regime, when $\delta \ll 1$; the mildly non-linear regime, when $\delta \sim 1$; and the highly non-linear regime, when $\delta \gg 1$. In the linear regime, the Equations 1.14 simplify to a second order partial differential equation for the density perturbation δ, given by

$$\frac{\partial^2 \delta}{\partial^2 t} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \rho_u \delta. \tag{1.17}$$

The solution to this equation can be written as the sum of a growing and decaying term

$$\delta(\mathbf{x}, t) = D_+(t)\delta(\mathbf{x}, t_0) + D_-(t)\delta(\mathbf{x}, t_0). \tag{1.18}$$

The part with $D_+(t)$ as the coefficient represents the growing mode solution, and the part with the coefficient $D_-(t)$ is the decaying mode solution. The decaying mode solution becomes negligible with time. The parameter $D_+(t)$ is the linear growth factor that depends on the total matter and energy content of the Universe via the relation

$$D(z) = \frac{5\Omega_m H_0^2}{2} H(z) \int_z^\infty \frac{1+z'}{H^3(z')} dz', \tag{1.19}$$

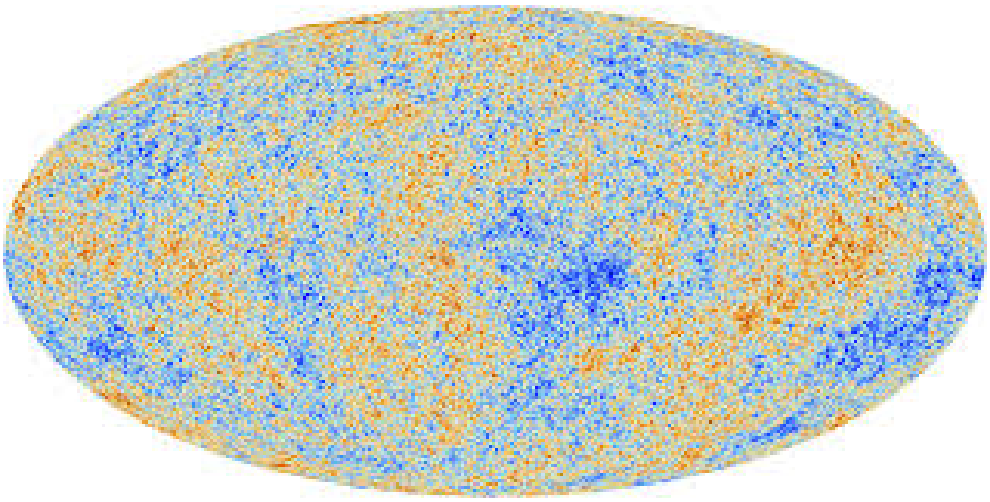


Figure 1.1 Latest measurement of the temperature anisotropies in the Cosmic Microwave Background as measured by the Planck satellite. The orange/red spots are locations where the temperature exceeds the average background temperature. The blue spots, are locations where the temperature is less than the global average. The color intensity is proportional to the excess/deficit temperature. Figure courtesy Planck Collaboration et al. (2013b).

where $H(z)$ is the Hubble factor at redshift z. The quantity $D_+(t)$ is of crucial importance, since it quantifies the growth of structure in the Universe. In most cases, its value grows with time. Note that the above expression is limited to a Universe filled with matter and cosmological constant, and neglects contributions from radiation and generic forms of dark energy.
The linear perturbation theory provides an exact description of the structure formation scenario as long as $\delta \ll 1$. Once $\delta \sim 1$, the linear theory approximations break down and is no long valid in these regimes.

1.2.3 The nature of initial conditions

As the structures in the Universe form from the originally tiny perturbations, it is of primary importance to characterize the nature of these initial perturbations. This is because the outcome of structure evolution is almost entirely dictated by the characteristics of the primordial perturbations. The nature of the early spatial perturbations is characterized as a Gaussian random field (Adler 1981; Bardeen et al. 1986; Heavens & Peacock 1988). The following reasons provide a compelling argument in favor of this hypothesis.

Inflation. Although there is no direct evidence, it is believed that the Universe underwent a phase of rapid inflation roughly 10^{-35}s after its birth. The "seeds" for the structures in the Universe are planted during the inflationary era in the form of spontaneous quantum fluctuations of the high energy pseudo-vacuum state that the

Universe was in. In the simplest of inflationary theories, these fluctuations are Gaussian distributed and therefore characterized as a Gaussian random field, with a scale invariant power spectrum of the form $P(k) = k^n$ (Guth & Pi 1982). The case when $n = 1$ is also predicted independent of the inflationary scenario (Harrison 1970; Peebles & Yu 1970; Zeldovich 1972). Somewhat more realistic inflationary scenarios predict that n deviates from unity, which agrees with measurements of $n \sim 0.96$ (Planck Collaboration et al. 2015).

Cosmic Microwave background. The temperature anisotropies in the Cosmic Microwave follow the anisotropies in the density perturbations directly, and offer a peek into the characteristics of the primordial density fluctuations. The measurement of the temperature fluctuations in the cosmic microwave background, suggests that its character is that of a homogeneous and isotropic Gaussian random field to high accuracy (Planck Collaboration et al. 2015; Smoot et al. 1992; Bennett et al. 2003; Spergel et al. 2007; Komatsu et al. 2010). Figure 1.1 shows the temperature anisotropies in the Cosmic Microwave Background radiation as measured by experiments like Planck (Planck Collaboration et al. 2015).

The Central Limit Theorem. The Central Limit Theorem, or CLT, states that a sum of many *independent and identically distributed* (i.i.d.) random variables will tend towards a normal distribution. The rapid expansion in the inflationary phase in the early Universe is most probably a result of an interaction of multiple scalar fields (Linde 1982a,b). In such a scenario, the observed quantum fluctuations are a superposition of multiple random fields, each corresponding to a given scalar field. Invoking CLT in such a case means that the net observed fluctuation, which is a superposition of multiple fields, follows a Gaussian profile.

1.2.4 Gaussian Random fields

A *random variable* is a variable whose value is subject to probabilistic variations. Random variables model the outcome of *stochastic processes*, the values of which vary randomly over the domain of the variable. The values of a random variables are not fixed, but rather drawn from a *distribution function*. The most common is the *Gaussian distribution function*, given by

$$f(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{1.20}$$

where, σ is the standard deviation of the variable. We implicitly assume that the mean or expectation of the variable is zero.

A generalization of a random variable is a *random field*, which is a collection of random variables. A random field is a *Gaussian random field* if the set of constituent random variables are all drawn from Gaussian distributions.

A random field can be specified by the m-point joint distribution function

$$P[f_1, \ldots, f_m] df_1 \ldots df_m. \tag{1.21}$$

The m-point joint probability distribution function for a Gaussian random field is a multi-variate Gaussian, given by

$$P\left[f_1,\ldots,f_m\right]df_1\ldots df_m \;=\; \frac{1}{(2\pi)^N(detM)^{1/2}}\cdot\exp\left(-\frac{\sum\Delta f_i(M^{-1})_{ij}\Delta f_j}{2}\right)df_1\ldots df_m \tag{1.22}$$

The equation is in the normalized form, such that the integral of $P\left[f_1,\ldots,f_m\right]df_1\ldots df_m$, over all $f\in\mathbb{R}^N$, is equal to 1. In the above expression,

$$\begin{aligned}\Delta f_i &= f_i-\langle f_i\rangle\\ M_{ij} &= \langle\Delta f_i\Delta f_j\rangle\end{aligned} \tag{1.23}$$

The matrix M^{-1} is the inverse of the $m\times m$ covariance matrix M_{ij}, in which the angle bracket denotes the ensemble average of the product, over the 2-point probability distribution function. In effect, M is the generalization of the variance of the 1-point normal distribution. Indeed, $M=[\sigma_0^2]$ for the case $m=1$.

1.2.5 Properties of Gaussian random fields: correlation function and power spectrum

Equation (1.22) is an expression of the fundamental property of Gaussian fields that they are fully specified by the second order moment, via the autocorrelation function $\xi(\mathbf{r})$, encoded through the covariance matrix. The correlation function expresses the correlation between the density values at any two points $\mathbf{r_1}$ and $\mathbf{r_2}$, such that

$$\mathbf{r}=\mathbf{r_1}-\mathbf{r_2}. \tag{1.24}$$

If the field is homogeneous and isotropic, the correlation function becomes a function only of the absolute distance between the points $r=|\mathbf{r}|$, such that

$$\xi(r)\;=\;\xi(|\mathbf{r}|)\;\equiv\;\langle f(\mathbf{x})f(\mathbf{x}+\mathbf{r})\rangle\,. \tag{1.25}$$

In other words, the entries in the matrix are the values of the *autocorrelation function* for the distance between the points: $M_{ij}=\xi(r_{ij})$, with $r_{ij}=\|\mathbf{x}_i-\mathbf{x}_j\|$.

To appreciate the contribution from different scales, the structure of a Gaussian field is more transparently characterized by the power spectrum $P(k)$. It is the Fourier transform of the autocorrelation function, specifying the mean square of the fluctuations of the Fourier components $\hat{f}(\mathbf{k})$ of the field $f(\mathbf{x})$,

$$f(\mathbf{x})=\int\frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}\,\hat{f}(\mathbf{k})\,e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{1.26}$$

via the relation

$$\langle \hat{f}(\mathbf{k}) \hat{f}(\mathbf{k}') \rangle \;=\; (2\pi)^{3/2}\, P(k)\, \delta_D(\mathbf{k} - \mathbf{k}')\,, \tag{1.27}$$

where $\delta_D(\mathbf{k})$ is the Dirac delta function. This implies that the knowledge of the power spectrum alone is sufficient to fully characterize a Gaussian random field. Below we give examples of two cosmologically relevant power spectra.

Power-law power spectrum. The power-law power spectrum is a generic class of spectra, specified by the spectral index n

$$P(k) \propto k^n. \tag{1.28}$$

The case when $n = 1$ is the scale invariant spectrum. Known commonly as the Harrison-Zel'dovich-Peebles spectrum, it is considered to be the natural choice for the primordial power spectrum (Dunkley et al. 2009; Komatsu et al. 2011; Planck Collaboration et al. 2015). This is to avoid a power spectrum with large rises either at large wavenumbers ($n > 1$) or small wavenumbers ($n < 1$). In these cases, δ could exceed unity and nonlinear collapsed structures (e.g. primordial black holes) could form in the ultra-early universe (van de Weygaert & Bond 2008). The measured spectrum of the primordial perturbations is very close to it, $n \sim 0.96$ (Dunkley et al. 2009; Komatsu et al. 2011; Planck Collaboration et al. 2015). It is worthwhile noting here that certain inflationary theories also predict $n \sim 1$ for the primordial power spectrum (Guth & Pi 1982).

LCDM power spectrum. The LCDM power spectrum stems from the standard concordance model of cosmology. It fits the measured power spectrum of the cosmic microwave background as well as the power spectrum measured in the nearby large scale Universe to high accuracy. It is given by (Eisenstein & Hu 1999; Hu & Eisenstein 1999)

$$P(k) \propto T^2(k) P_0(k), \tag{1.29}$$

where $P_0(k)$ is the primordial power spectrum and $T(k)$ is the transfer function, which is an expression of physical processes acting on the fluctuations as they enter the horizon. Its shape can be inferred by evaluating the evolving processes, through the Boltzmann equation. A good numerical fit is given by (Eisenstein & Hu 1999; Hu & Eisenstein 1999)

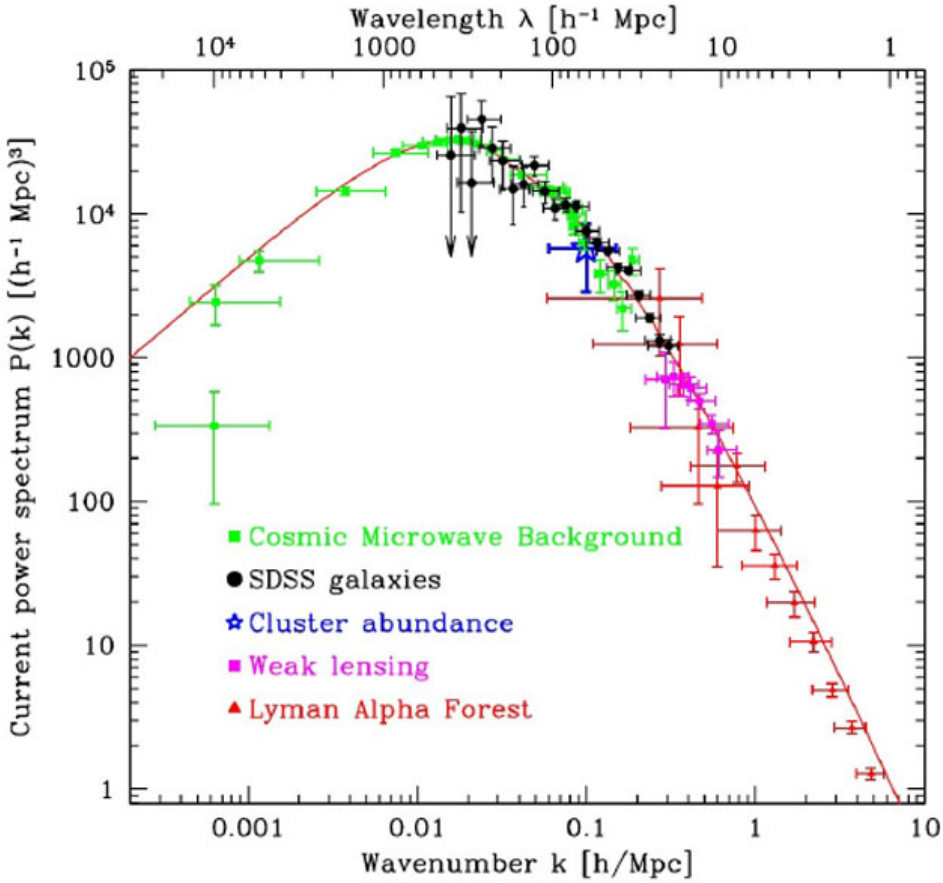


Figure 1.2 The theoretical power spectrum as predicted by the concordance cosmology. Superimposed are also the observationally measured power spectra over approximately four decades of wave number. Figure reproduced from Tegmark et al. (2004).

$$P_{CDM}(k) \propto \frac{k^n}{\left[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right]^{1/2}} \times \frac{\left[ln(1+2.34q)\right]^2}{(2.34q)^2}, \tag{1.30}$$

$$q = k/\Gamma,$$

$$\Gamma = \Omega_m h \exp\left\{-\Omega_b - \frac{\Omega_b}{\Omega_m}\right\}.$$

In the above expression, Ω_m and Ω_b are the total matter density and baryonic matter density respectively. Γ is referred to as the shape parameter. Plugging in the most recent observed values of Ω_b and Ω_m, $\Gamma \sim 0.21$.

The LCDM power spectrum is usually normalized by the amplitude of fluctuations in the Cosmic Microwave Background. Figure 1.2 presents the theoretically predicted power spectrum from LCDM cosmology (in red). Superimposed in this figure are also measurements using cosmic microwave background (CMB) anisotropies, galaxy large scale structure, weak lensing of galaxy shapes, and the Lyman alpha forest, in order of decreasing comoving wavelength (Tegmark et al. 2004). In addition, there is a single data point for galaxy clusters, whose current space density measures the amplitude of the power spectrum on $8h^{-1}$Mpc scales (White et al. 1993). One can see that the modern CMB and LSS data agree over a substantial region of overlap with the theoretical power spectrum, giving us reasonable confidence in the correctness of the model.

1.2.6 The non-linear regime

When $\delta \sim 1$, linear theory is no longer adequate in explaining the evolution of structures due to non-linear coupling between different modes. Structure growth in the non-linear regime has been understood in great details through elementary, yet elegant, models like the *spherical collapse* model (Gunn & Gott 1972) and the *ellipsoidal collapse* model (Icke 1973; White & Silk 1979; Peebles 1980).

In the spherical collapse model, one examines the non-linear evolution of a spherically symmetric density peak. Under gravitational amplification, this primordial density peak turns around and collapses, when the linearly-extrapolated primordial density reaches a critical density f_c. The peaks in the primordial field that have collapsed into collapsed objects at any time t (and hence the redshift z) are those for whom the overdensity linearly extrapolated to the present epoch satisfies

$$\delta_L(\mathbf{x}, t|R) > f_c(z)/D(z). \tag{1.31}$$

Here, δ_L is the linearly extrapolated overdensity, and $D(z)$ is the structure factor (see Equation 1.19).

While the spherical collapse model is highly instructive in informing about the mechanism of non-linear growth of structures, it is nevertheless a very idealized scenario. In realistic circumstances, the primordial density perturbations are never spherical, nor isolated (Bardeen et al. 1986). The overdensity in realistic circumstances collapse anisotropically.

Historically, anisotropic collapse has been modeled by studying the collapse of homogeneous ellipsoids (Icke 1973; White & Silk 1979; Peebles 1980). Icke (1973) showed that the collapse along the shortest axis will occur more rapidly than that of the spherical equivalent; the collapse along the medium axis will be comparable to the spherical value, while the full collapse along all three axes will be slower than the spherical counterpart. The extension to a cosmological setting, through inclusion of the relation of exterior tidal forces to the linear deformation tensor of the interior

was worked out by Bond & Myers (1996a).

The spherical collapse model is useful in gaining understanding about the conditions when a genuinely recognizable cosmic object may form. In the same vein, the ellipsoidal collapse model provides a basic understanding about the anisotropic nature of the structural collapse. However, both these simple models cannot account for the hierarchical nature of structures observed in the Universe.

1.2.7 Hierarchical growth of structures

In the hierarchical structure formation scenarios, structures form *bottom-up*. The first structures to form are small compact objects. These objects merge together to form larger and larger structures. Extended features that are still in the process of collapsing, or collapsed features that have not fully virialized yet, contain a large amount of smaller scale substructures at higher density. These substructures are a telling example of the hierarchical structure formation process in the Universe. This *hierarchical nature*, together with *anisotropy* is perhaps the most significant property of the cosmic mass distribution (Press & Schechter 1974; Cole & Lacey 1996; Lindner et al. 1995; Houlahan & Scalo 1992; Abel et al. 1998).

Observationally, the indications of hierarchical structure formation process can be recognized in the megaparsec scale Universe. On these scales, one may recognize large unrelaxed filamentary as well as sheet-like structures containing a multitude of collapsed and virialized structures like rich clusters of galaxies as well as smaller galaxy groups. On smaller scales, within the galaxy groups, the larger galaxies are accompanied by smaller dwarf galaxies. Imprints of hierarchical structure formation are visible even at the scales of fully collapsed objects like clusters of galaxies. A telling example is the Coma cluster which has a heavy group falling in (Neumann et al. 2003). At the scales of galaxies, the imprints of hierarchical structure formation process is manifested through the presence of streams in their dark haloes, which most likely are remnants of infalling dwarf galaxies (Helmi et al. 1999; Helmi & White 1999).

Computer simulations also show that larger and larger structures form due to mergers and accretion of smaller structures that have formed at earlier times. Figure 1.3 shows zoom-ins at various scales in the state-of-the-art *Millennium II* simulation (Boylan-Kolchin et al. 2009). At successive scales, the structure is marked by anisotropic filamentary and sheet-like structures, that contain clusters of galaxies at smaller scales.

1.2.8 Number distribution of collapsed objects : (extended) Press-Schechter formalism

In the hierarchical structure formation scenario, it is possible to statistically account for the number and mass distribution of gravitationally collapsed objects using the Press-Schechter formalism (Press & Schechter 1974). It is based on simple geometric considerations. The two key aspects are that the mass density fluctuations are Gaussian in nature, and the overdensities collapse spherically (Gunn & Gott 1972), meaning there is no preferred axis of collapse. It predicts that the number density $n(M)$ of objects with mass M is given by (Press & Schechter 1974; van de Weygaert

& Bond 2008)

$$\begin{aligned} n(M)\,dM &= \sqrt{\frac{2}{\pi}}\frac{\rho_u}{M^2}\nu(M)\,e^{-\frac{\nu^2(M)}{2}}\,|\frac{dln\sigma(M)}{dlnM}|, \\ \nu(M) &= \frac{f_c}{\sigma(M)}. \end{aligned} \tag{1.32}$$

In the above expression ρ_u is the universal background density, $\sigma(M)$ is the rms fluctuation at mass scale M, and is basically the power spectrum. This is related to to the length scale R through the expression

$$M \propto R^3. \tag{1.33}$$

$\nu(M)$ is the dimensionless threshold decided by the collapse barrier f_c and the rms mass fluctuation $\sigma(M)$.

More recently, the formalism has been extended to a more detailed and accurate analysis based on the excursion set formalism (Peacock & Heavens 1990; Bond et al. 1991; Sheth 1998; Sheth & Tormen 2002), and is commonly known as the *extended* Press-Schechter formalism.

The PressSchechter formalism can also be applied to calculate the number distribution of voids. The cosmic voids are defined as regions whose density is less than some critical value $\delta_\nu \leq 0$ or, alternatively, as regions for which the three eigenvalues of the tidal tensor (Doroshkevich 1970; Bardeen et al. 1986) lie below some critical value $\lambda_\nu \leq 0$ (Sheth & van de Weygaert 2004; Kamionkowski et al. 2009; Lam et al. 2010; Song & Lee 2009; Jennings et al. 2013; Chan et al. 2014). An important aspect in the calculation of the probability function of voids is the over-counting of voids located inside collapsing regions. This *voids-in-clouds* problem (Sheth & van de Weygaert 2004), can be solved within the excursion set theory by studying a two-barrier problem: δ_c for halos and δ_v for voids. A considerable advantage of studying voids and the void probability function is that the shape, size and number density of voids is extremely sensitive to the underlying dark energy model, and hence can be a potentially powerful tool for probing the nature of dark energy (Bonometto 1995; Park & Lee 2007b; Kamionkowski et al. 2009; Lavaux & Wandelt 2010; Bos et al. 2012).

A very crucial point to note is that the (extended) Press-Schechter formalism relates to the singularity structure of the density field. In particular, it identifies cosmological objects through the critical points, namely the maxima, saddles and minima, of the density field. It computes the collapse and the corresponding number density of objects by tracking the upward crossing of singularities through the density barrier defined by f_c. By invoking the singularity structure of the density field, the above formalisms are intimately connected to the topological structure of the field.

1.2.9 Numerical simulations

Numerical simulations have become the standard tool to understand the formation and evolution of structures in the highly non-linear regimes. It has become possi-

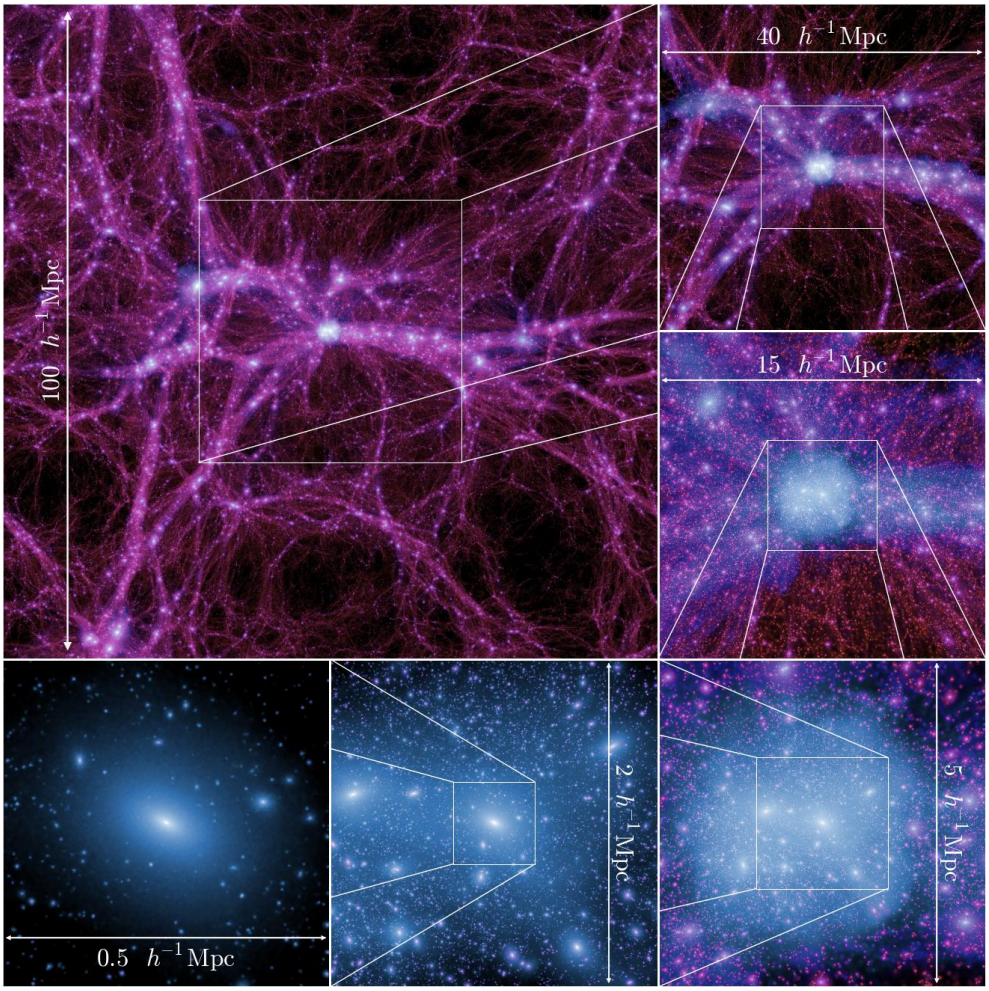


Figure 1.3 A sequential zoom through the Millennium-II Simulation depicting the hierarchical nature of structures in the Universe. The simulation reveals the anisotropic and multiscale character of the hierarchically evolving mass distribution. The large image (upper left) is a $15h^{-1}$Mpc thick slice through the full $100h^{-1}$Mpc simulation box at redshift zero, centered on the most massive halo in the simulation. Starting from the upper right and moving clockwise, subsequent panels zoom into the cluster region and show slices that are 40, 15, 5, 2, and $0.5h^{-1}$Mpc on a side (with thicknesses of 10, 6, 5, 2, and 0.5 h^{-1}Mpc). Even at $0.5h^{-1}$Mpc, which is approximately 1/10th the diameter of the halo, a rich variety of substructure is visible. Figure and text courtesy (Boylan-Kolchin et al. 2009).

ble to track and focus on the hierarchical growth of even individual structures, and visualize the process of hierarchical evolution of structures through numerical simulations. N-body simulations, in principle involve the integration of 6N ordinary differential equations defining the particle motions in Newtonian gravity. The first computer simulation in astrophysics was implemented as early as 1960 (von Ho-

erner 1960). The earliest N-body simulations in a cosmological context were those of Aarseth and collaborators in the 1970's and 1980s (Aarseth 1972, 1973; Aarseth et al. 1979; Aarseth & Dekel 1983; Aarseth 2011).

The majority of simulations aimed at understanding the distribution and evolution of matter at large scales in the Universe are dark matter simulations. To first approximation, this is justified because the major share of matter in the Universe is dark matter. This means that dark matter is the gravitationally dominant mass component in the structure formation process. As a result, it provides us with an accurate outline of the full cosmic mass distribution. It is considerably more challenging to include baryonic physics in simulations. This is because baryonic physics is rather poorly understood and involves a multitude of processes like star formation and radiative transfer, which make it considerably more expensive computationally.

The Millennium simulation (Springel 2005) is a state of the art computer simulation that uses 2160^3 particles (more than 10 million) to trace the evolution of the matter distribution in a cubic region of the Universe $500h^{-1}$Mpc on a side. Figure 1.3 shows the hierarchy of structures as seen in the Millennium II Simulation (Figure courtesy Boylan-Kolchin et al. (2009)).

Due to the continuous and almost exponential increase in computational power, we have seen some impressive developments when it concerns simulating the dark matter and the baryonic matter together. Most recently, the Illustris project (Vogelsberger et al. 2014) has successfully simulated in extraordinary detail, the full array of dark matter and baryonic physics. This is in a box of side $100h^{-1}$Mpc. The size of the box and the resolution is optimal to account for the formation of large scale structures, while simultaneously being able to simulate baryonic physics as well as black hole formation.

1.3 The Cosmic web

While we think that the cosmic matter distribution is uniform on large Gigaparsec scales, it has an intricate web-like pattern on smaller distances. At scales of few to hundreds of megaparsecs, [1] the universe has a web-like appearance, commonly known as the *Cosmic Web* (Bond et al. 1996). The cosmic web is composed of huge, virialized blobs called *clusters*, containing thousands of galaxies, as well as huge amounts of dark matter and high temperature gas. These objects, due to their high mass and luminosity, represent the most prominent feature of the cosmic web. One such typical example, the Coma cluster, can be seen at the center of the CfA2 Great Wall in the lower wedge in the top quadrant of Figure 1.4. Out of these virialized clusters, spread out gigantic *filaments*. One can notice these filaments in all the panels of Figure 1.4. Notice, for example, the thready structures coming out of the Coma cluster. These filaments, also containing thousands of galaxies, dark matter and gas, serve as transport channels for mass flow in the Universe. The cosmic web also consists of huge sheet-like objects commonly known as *walls*. The clusters, filaments

[1] A parsec is the standard unit of measurement of distances in the cosmos. A parsec is 3.26 times the *light-year*, the distance light covers in a year. A megaparsec is a million parsecs, the typical scale of measurement of size of the large scale structures in the universe.

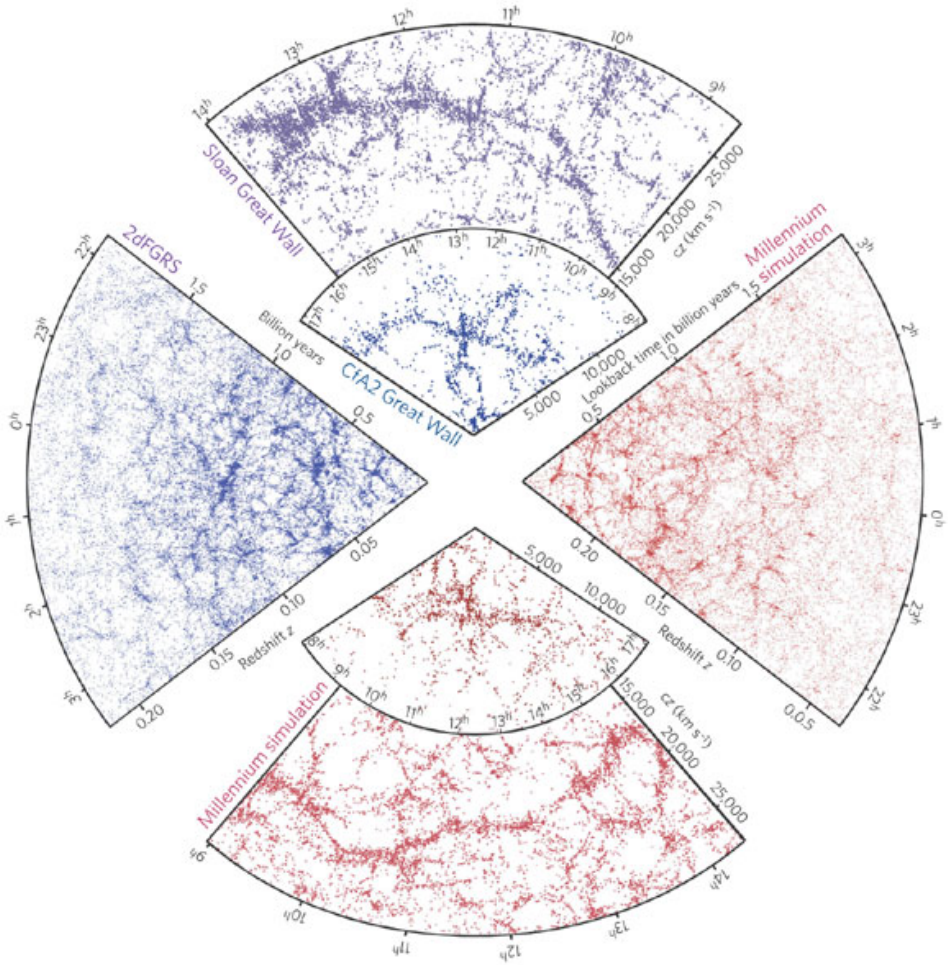


Figure 1.4 The cosmic web. The galaxy distribution obtained from spectroscopic redshift surveys and from mock catalogues constructed from cosmological simulations. The small slice at the top shows the CfA2 Great Wall, with the Coma cluster at the center. The CfA2 great wall contains around 2000 galaxies, and stretches across $230h^{-1}$Mpc. Drawn to the same scale is a small section of the Sloan Digital Sky Survey, in which an even larger Sloan Great Wall has been identified. This is one of the largest observed structures in the Universe, containing over 10,000 galaxies and stretching over more than $420h^{-1}$Mpc. The wedge on the left shows one-half of the 2dFGRS survey, which determined distances to more than 220,000 galaxies in the southern sky out to a redshift of 0.25. The SDSS has a similar depth but a larger solid angle and currently includes over a million observed redshifts in the northern sky. At the bottom and on the right are shown the mock galaxy surveys, constructed from the "Millennium simulation. These are constructed such that the survey geometries and magnitude limits match those of the observational catalogs.

and walls of the web are intertwined around vast near-empty *cosmic voids*. As can be seen in all the panels in Figure 1.4, both observations and simulations present the evidence that cosmic voids are the dominating components in terms of volume occupancy (also see, e.g., Cautun et al. (2014)).

That the cosmic web indeed exists has been confirmed by a host of galaxy redshift surveys like the CfA and CfA2 surveys (de Lapparent et al. 1986; Huchra & Geller 1982), the 2dFGRS survey (Colless et al. 2003), the SDSS survey (Abazajian et al. 2003) and the 2MRS survey (Huchra et al. 2005; Skrutskie et al. 2006). Massive megaparsec to gigaparsec scale computer simulations also show the existence of the Cosmic web (Springel 2008). Figure 1.4 presents an impression of the galaxy distribution obtained from spectroscopic redshift surveys, as well as from mock catalogues constructed from cosmological simulations (Figure courtesy Springel et al. (2006)). The small slice at the top shows the CfA2 Great Wall, with the Coma cluster at the center The CfA2 great wall contains around 2000 galaxies, and stretches across $230h^{-1}$Mpc. Drawn to the same scale is a small section of the Sloan Digital Sky Survey, in which an even larger Sloan Great Wall has been identified. This is one of the largest observed structures in the Universe, containing over 10,000 galaxies and stretching over more than $420h^{-1}$Mpc. The wedge on the left shows one-half of the 2dFGRS survey, which determined distances to more than 220,000 galaxies in the southern sky out to a redshift of 0.25. The SDSS has a similar depth but a larger solid angle and currently includes over a million observed redshifts in the northern sky. At the bottom and on the right are shown the mock galaxy surveys, constructed from the "Millennium simulation. These are constructed such that the survey geometries and magnitude limits match those of the observational catalogs. One can notice the presence of similar kind of structures in all the four wedges that concern both observational data and simulations.

The theoretical understanding of the nature of the emergent web from the primordial fluctuation field is now well developed (a brief description of the theory of formation and evolution of structures in the cosmos can be found in Section 1.2). However, an in-depth analysis is hindered by a multitude of factors. Chief among these are the lack of symmetry in the structures, strong non-local influences, and the hierarchical nature of the gravitational clustering, due to which many scales are simultaneously relevant. The structures in the web also exhibit complex connectivity, wherein the structures of lower dimension like filaments are embedded within structures of higher dimensions like walls. In addition, one also sees the hierarchical buildup of structures in the web, wherein many small structures connect together to form progressively larger structures. The complex connectivity, in combination with the hierarchical buildup of structures, makes it an extremely challenging task to segregate the components of the web, and study their properties.

1.3.1 Zel'dovich approximation

The Zel'dovich approximation (Zel'dovich 1970) assumes a seminal role in understanding the formation of web-like anisotropic structures that pervades the Universe. It is a Lagrangian formulation, in which the frame of reference moves along with the mass element of interest. It looks at the displacement and corresponding

Component	λ_1	λ_2	λ_3
Clusters	+	+	+
Filaments	+	+	-
Walls	+	-	-
Voids	-	-	-

Table 1.2 The morphology of a mass element according to the Zeldovich formalism. The quantities $\lambda_1 > \lambda_2 > \lambda_3$ denote the eigenvalues of the deformation tensor. Their signs determines the morphology of the element.

deformation of mass elements at the initial Lagrangian location. It can be shown that the position $\mathbf{x}(t)$ at some time t depends on the initial coordinate q and a displacement term via

$$\mathbf{x}(t) = \mathbf{q} - D_+(t)\vec{\nabla}_q \Psi(\mathbf{q}). \tag{1.34}$$

Here $D(t)$ is the linear growth factor, and $\Psi(\mathbf{q})$ is related to the initial peculiar gravitational potential $\phi_0(\mathbf{q})$ via the relation

$$\Psi(\mathbf{q}) = \frac{2}{3\Omega_0 H_0^2}\phi_0(\mathbf{q}). \tag{1.35}$$

The peculiar gravitational potential $\phi_0(\mathbf{q})$ can, in turn, be related to the perturbations in the initial density field.

Using this prescription, we can evaluate the evolution of the initial mass element $\bar{\rho}\mathrm{d}^3\mathbf{q}$ to a the element $\rho(t)\mathrm{d}^3\mathbf{x}$, at a later time t. The mass within the volume element in question is conserved at all times, such that

$$\rho_0 \mathrm{d}^3\mathbf{q} = \rho(t)\mathrm{d}^3\mathbf{x}. \tag{1.36}$$

This in turn gives the evolution of the density field

$$\rho(\mathbf{x}) = \frac{\rho_0}{[1 - D\lambda_1(\mathbf{q})][1 - D\lambda_2(\mathbf{q})][1 - D\lambda_3(\mathbf{q})]}, \tag{1.37}$$

where, $(\lambda_1 > \lambda_2 > \lambda_3)$ are the eigenvalues of the deformation tensor

$$\Psi_{ij} = \frac{\partial^2 \Psi}{\partial q_i \partial q_j}. \tag{1.38}$$

Equation 1.37 provides a recipe for the anisotropic collapse of matter into pancakes, filaments and clusters. The final morphology is determined by the magnitude as well as the sign of the eigenvalues, which is presented in Table 1.2. A positive eigenvalue represents a compression along that direction, while a negative eigenvalue implies an expansion. When $\lambda_1 > 0$, it implies that $(1 - D\lambda_1) \to 0$ at some

time t corresponding to

$$D(t) = 1/\lambda,$$

and a collapse along that axis results in a sheet. Similarly, for $(\lambda_2, \lambda_3) > 0$ implies a collapse along the second and the third axis, giving rise to a filament and a cluster respectively. The condition $(\lambda_1 > \lambda_2 > \lambda_3)$ also points to the fact that the collapse and formation of anisotropic structures follows a well-defined sequence – from pancakes, to filaments, to clusters.

The Zel'dovich formalism, however, suffers from the problem that it is a ballistic motion, that results in extended shapes of collapsed objects. This is an artifact of the formalism. To remedy this, one needs to *stick* these particles in the dense region, which is taken care of by introducing an additional viscosity term. This extension of the Zel'dovich approximation through the approach of *adhesion* was first advocated by Kofman & Shandarin (1988), Gurbatov et al. (1989) and Kofman et al. (1990). It has received renewed attention in the recent years due to the work of Hidding et al. (2014, 2015).

A more comprehensive analytical description of the hierarchical structure formation, embedding the anisotropic collapse into hierarchical clustering scenario, has been developed by Bond and Myers (Bond & Myers 1996a,b,c), which is known as the *peak patch* formalism. Assuming that the initial density field is Gaussian, this formalism shows that it is enough to know the tidal field in a few well chosen locations which are the nodes, or clusters, to delineate the patterns of the web in that region (Bond et al. 1996). It highlights the importance that virialized clusters play in determining the overall connectivity and shape of the web.

1.3.2 The importance of the Cosmic Web

Over the last few decades, the importance of determining the properties of the cosmic web has become increasingly important. The patterns in the web have significant implications for the understanding of the growth of structures in the Universe, as well as for the understanding of formation and evolution of galaxies. Off late, the idea that the Cosmic Web may provide an independent and additional probe for testing deviations of primordial density perturbations from Gaussian initial conditions is also being recognized widely (Verde et al. 2000; Bartolo et al. 2005; Hikage et al. 2006, 2008; McDonald 2008).

The different components of the Cosmic Web can provide independent probes for constraining cosmological models, as well as estimates on the amount of dark matter and dark energy in the Universe. For example, the structure growth factor, measured through observations of clustering in the cosmic mass distribution, offers a window into the role that dark energy plays in governing the dynamics of the Universe in the late epochs (Guzzo 1996; Linder & Jenkins 2003; Seljak et al. 2002; Lavaux & Wandelt 2010; Angulo et al. 2012; Sutter et al. 2014a). The estimation of the total amount of dark matter in the Universe, can be probed through gravitational lensing of distant galaxies by the clusters and filamentary structures in the Cosmic Web (Dietrich et al. 2012).

The measurement of the evolution and influence of dark energy has a potential probe through the cosmic voids (for a review, see van de Weygaert & Platen 2011).

The underdense nature of cosmic voids makes them more sensitive to the influence of dark energy than any other component of the web. Probes of the influence of dark energy on the voids include determining the evolution of dark energy through measuring the shape and size of cosmic voids (Park & Lee 2007b; Lavaux & Wandelt 2010; Bos et al. 2012). Cosmic voids also acquire importance in the modified gravity models, as the shape and size of voids is more sensitive to the details of the model in comparison to clusters, filaments or walls (Clampitt et al. 2012; Terukina & Yamamoto 2012; Hellwing et al. 2014). This is because in these models, the fifth force or the modified gravity term is heavily screened in the high density regions.

A tool aimed at studying the nature of cosmic voids is the void abundance or the density dependent frequency of cosmic voids (Jennings et al. 2013; Chan et al. 2014; Pisani et al. 2015). A considerable advantage of studying voids and the void probability function is that the shape, size and number density of voids is extremely sensitive to the underlying dark energy model, and hence can be a potentially powerful tool for probing the nature of dark energy (Bonometto 1995; Park & Lee 2007b; Kamionkowski et al. 2009).

The Cosmic Web is also of key importance from the view of understanding about the formation, evolution and properties of galaxies. The formation, evolution and properties of galaxies has been shown to depend on the morphology as well as density of the part of the web where it resides in (Cautun et al. 2014).

The anisotropic nature of matter distribution in the web also generates tidal fields, which is known to be the responsible for generation of angular momentum in galaxies (Hoyle 1951a; Doroshkevich 1970; Efstathiou & Jones 1979; Jones & Efstathiou 1979). This has been investigated in numerical simulations to show that angular momentum as well as the shape of dark matter haloes is strongly correlated with the orientation of the host large scale structures (Aragón-Calvo et al. 2007a; Hahn et al. 2007; Libeskind et al. 2012; Aragon-Calvo & Yang 2014). That the orientation of angular momentum of galaxies is strongly correlated to the morphology of the host large scale structure, in particular the filaments, is also evident in the observations of alignment of spin axis with respect to the large scale filament they reside in (Jones et al. 2010; Tempel et al. 2013; Tempel & Libeskind 2013).

Another interesting environment with respect to galaxy evolution is voids. The void galaxies provide a unique perspective regarding galaxy formation and evolution, because of the relatively pristine environment uncontaminated by complex processes like mergers and tidal stripping, that have more significant effect in denser regions (van de Weygaert & Platen 2011; van de Weygaert et al. 2011; Kreckel et al. 2012).

1.4 Characterization of the cosmic web

The effort to describe and characterize the mass distribution in the Universe has been ongoing for decades now. A quantitative characterization of the complex and rich spatial patterns in the Universe is extremely important in view of comparing and validating various models with respect to the actual mass distribution in the Universe. The methods employed towards this end can broadly be classified into

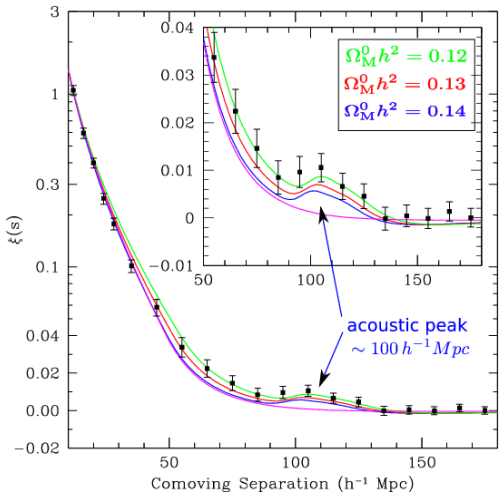


Figure 1.5 The large scale redshift-space correlation function of the SDSS LRG (luminous red galaxies) sample. The black dots represent the correlation function calculated from observed galaxy distribution. The magenta line shows a pure CDM model ($\Omega_m h^2 = 0.105$), meaning there are no baryons. Characteristically, it lacks the baryonic acoustic peak. The green, red and blue lines represent models with a baryon fractional density of $\Omega_b h^2 = 0.024$ and a CDM density of, respectively, $\Omega_m h^2 = 0.12, 0.13, 0.14$. On the smaller scales, the correlation function approximates a power law. On progressively larger scales the correlation function drops down near zero, signifying isotropy. The bump at $100h^{-1}$Mpc corresponds to the baryonic acoustic peak. Figure courtesy Eisenstein et al. (2005).

three categories – statistical, geometric or morphological, and topological. Below we present a short account of these techniques and discuss a few relevant results wherever possible.

1.4.1 Statistical quantification of the web

Correlation functions. Measuring the n-point correlation function, more specifically the 2- point correlation function, has been the mainstay of statistical analysis in cosmology (Peebles 1980; Davis & Peebles 1983; Hamilton 1993; Szapudi & Szalay 1993a,b) (for an excellent and comprehensive textbook on the statistical characterization of the large scale mass distribution, see Martínez & Saar 2002).

Given a random point (galaxy) at location $\vec{x_1}$, in an infinitesimal volume dV_1, the 2-point correlation function describes the probability that another galaxy will be found at a location $\vec{x_2}$, in an infinitesimal volume dV_2. It is given by

$$dP_{\vec{x_1},\vec{x_2}} = \bar{n}^2 \left(1 + \xi(r_{12})\right) dV_1 dV_2, \tag{1.39}$$

where, $\bar{n}$ is the average number density of galaxies in the given survey volume, and $r_{1,2} = |\vec{x}_1 - \vec{x}_2|$. The continuous version of the 2-point correlation function is the *auto-correlation* function, which for a continuous density field describes the probability

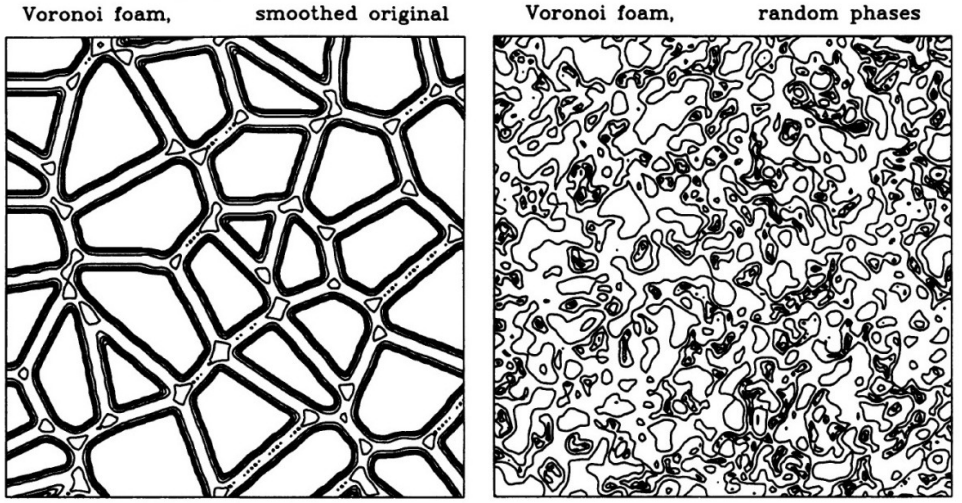


Figure 1.6 An illustration depicting the insensitivity of the correlation function to the phases of the constituent modes. On the left side is a structural pattern depicting a cellular structure. The right hand figure is produced by keeping the amplitudes of the constituent Fourier modes of the left-hand figure intact, while at the same time randomizing the phases of the Fourier modes. Since the correlation function is dependent only on the amplitude, both the patterns produce the same correlation function. Figure courtesy Szapudi & Szalay (1997).

that the density is ρ_1 at location $\vec{x_1}$ and ρ_2 at location $\vec{x_2}$

$$\xi(r_{12}) = \langle \rho(\vec{x_1})\rho(\vec{x_2}) \rangle. \tag{1.40}$$

Statistical isotropy and homogeneity, which is the fundamental assumption of the cosmological principle, implies that the auto-correlation function is a function only of the absolute distance between the locations (see Equation 1.25). Figure 1.5 presents the correlation function as measured from the distribution of luminous red galaxies (LRG) in the SDSS catalog (Peebles 1980; Eisenstein et al. 2005). The black dots represent the correlation function as measured from the observed galaxy distribution. The continuous lines are fits from various models. At small scales, the correlation function approximates a power law

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{1.41}$$

where, $\gamma = 1.8$ and $r_0 \simeq 5h^{-1}$Mpc. At progressively larger scales, the correlation drops down close to zero, signifying that the Universe is homogeneous on those scales (Peebles 1980; Martinez et al. 1990).

The correlation function has provided a wealth of statistical information about the nature of the matter distribution in the Universe. However, the auto correlation

function has no structural sensitivity to the geometric and morphological patterns of the mass distribution. This is because, the auto-correlation function is dependent only on the amplitude of the constituent Fourier modes, and provides no information about the phases of these modes. For example, it cannot distinguish between distributions with phase correlated Fourier modes versus distributions which have the phases of the Fourier modes randomly distributed. A revealing illustration of this is displayed in Figure 1.6. The left panel in the figure displays a pattern with a clear cellular structure. The pattern in the right panel is constructed from the pattern in the left panel by randomizing the phases of the constituent Fourier modes. Since the correlation function is dependent only on the amplitude, both the patterns produce the same correlation function. From this example, it is clear that the correlation function only provides limited or no amount of information on the complex geometric and morphological patterns of the cosmic web.

Higher order correlation functions, in particular the 3-point correlation function, have also been used to describe the statistics of matter distribution. However, the higher order correlation functions are limited in their usefulness because the errors in their measurement become prohibitively large with increasing order.

Count in cells and fractals. For both numerical models and analytic approximations of the cosmic mass distribution, it is necessary to characterize the clustering that develops in a quantitative manner. Conventionally, this is done by presenting the two-point correlation function $xi(r)$ for the mass distribution. However, by itself this does not fully characterize the distribution of points. An important alternative is to look at the distribution of counts in cells as a function of cell size (Jones 1992).

The *counts in cell* method represents another statistical approach, that is sensitive to the pattern of the mass distribution (Peebles 1980). The first analyses of galaxy sky maps were done by dividing the sky into cells and counting the cell occupancy (Bok 1934; Mowbray 1938; Rubin 1954; Limber 1954; Totsuji & Kihara 1969). It defines the probability of finding N galaxies in a volume V (White 1979). Particularly interesting is the *void probability function* $P_0(V)$

$$P_0(V, n_0) = exp(-an_0V), \tag{1.42}$$

that represents the probability that a volume V randomly selected in a point distribution with mean density n_0 will contain no galaxies (White 1979). White (1979) also point out that the scale a can be expanded in terms of the correlation functions of the point distribution, given by

$$a = 1 + \sum_{0}^{\infty}(-n_0)^{i-1}\xi_i \, dV_1 \, dV_2 \ldots \, dV_n. \tag{1.43}$$

where, ξ_i is the i-point correlation function of $(i-1)$ coordinates. Different clustering models have been proposed based on particular choices for the counts in cells (Coles & Jones 1991; Saslaw 2000; Borgani 1993).

There is a formalism for describing moments of cell counts that is commonly used when describing fractal point sets that was adopted as a clustering descriptor

of the cosmic mass distribution by Martinez et al. (1990). Possible ways of analyzing the statistical properties of point sets is through the possible scaling of the moments of the counts in cells (Jones 1992) or the scaling of moments of counts of neighbors (Martinez & Coles 1994).

The fractal model is of relevance only in the non-linear regimes of gravitational clustering. In the linear regimes, we do not expect any scaling, as the mass distribution follows from the primordial field. On large scales, the Universe is isotropic and homogeneous and hence lacking in any structure. Evidence from galaxy distributions as seen in the SDSS also suggest that the distribution of matter smooths out on a scale of around $100h^{-1}$Mpc (Groth & Peebles 1977, 1986; Peebles 1993). Important to note also is that the fractal description implies no particular underlying physical process: it is merely a statement of how moments of counts in cells behave as a function of cell size (Jones 1992).

1.4.2 Delineating the structures of the web

Over the last decade the view has emerged that a full analysis of the rich structures seen in the mass distribution in the Universe need to take into account the geometric and morphological aspects of the elements of the cosmic web. In particular, there has been an effort to detect and isolate the components of the cosmic web, namely clusters, filaments, walls and voids, with a view to study their characteristics in a more systematic way. In recent years, a number of methods attempting to detect and describe the structural patterns in the cosmic web have been developed. These methods have been immensely instructive in illuminating the connectivity features and the structural patterns of the web.

Studies carrying out an analysis of the global connectivity of cosmic web, by means of percolation analysis (Shandarin & Zeldovich 1983) were the first to do so with respect to the filaments in the galaxy distribution. An attempt to construct the minimal spanning tree (Barrow et al. 1985; Colberg 2007) of the structures in the cosmic web has also been instructive in studying the connectivity features of the cosmic web. Recently, the methodology has been extended to produce a full characterization of the cosmic filamentary structures in the web by the GAMA (Galaxy and Mass Assembly) team (Alpaslan et al. 2015). Percolation analysis and spanning trees, while relatively sensitive to the underlying structure, are still global quantities and fail to capture and describe the local variations in shape.

Within this, the statistical measures include the analysis of the global connectivity of Cosmic Web, by means of *percolation analysis* (Shandarin & Zeldovich 1983) and the *minimal spanning tree* (Barrow et al. 1985) of the galaxy distribution. Percolation analysis and spanning trees, while relatively sensitive to the underlying structure, are still global quantities and fail to capture and describe the local variations in shape.

Noteworthy also are the models and methodologies emanating from statistical and geometric considerations for delineating the cosmic filaments. Most prominent among them are the *candy model*, that uses a marked point process to delineate the filaments (Stoica et al. 2005; Tempel et al. 2013), and the recovery of filaments by recognizing them as the medial axis of a given point process (Genovese et al. 2010).

Morphological methods aimed at determining shapes locally considering the properties of the eigenvalues of the Hessian of a range of fields arising in cosmology have also been at the forefront in delineating the cosmic web. These involve the eigenvalues of the Hessian of density field (Aragón-Calvo et al. 2007b; Cautun et al. 2013) or those of the tidal tensor or the velocity shear tensor (Tempel et al. 2014). A variation of these methods further evaluates this in a multi-scale fashion by performing the procedure on a stack of images in the scale space, which are derived from the original image through convolution with an appropriate filter at a range of scales (Aragón-Calvo et al. 2010; Cautun et al. 2014).

Morphology based shape determination methods are closely related to the shape determination methods arising from topological considerations. Both the methods involve the identification of a morphological element by considering the properties of the eigenvalues of the Hessian of the density field for either extended objects (morphological methods), or that of a localized point in space (topological methods).

Besides being useful for delineating the structures in the web, topological methods are also important from the point of view of investigating the inherent structure and connectivity between the various structural elements that form the complex patterns that is the Cosmic Web. In the next section, we present a brief account of topology, with a focus on cosmological applications. This is the main theme of this thesis.

1.5 Topology of the mass distribution in the Universe

Topology (greek: "$\tau\acute{o}\pi o\varsigma$" (topos), meaning "space", and "$\lambda\acute{o}\gamma o\varsigma$" (logos) meaning "study") is the mathematical study of shapes. It is an area of mathematics concerned with the properties of space that are preserved under continuous deformations including stretching (compression) and bending, but not tearing or gluing. This includes the study of properties of spaces such as connectedness, continuity and boundary. Topology developed as a field of study out of geometry and set theory, through analysis of such concepts as space, dimension, and transformation.

Topology is a vast field, which includes many subfields. *General topology* establishes the foundational aspects of topology and investigates properties of topological spaces, as well as concepts inherent to topological spaces. *Algebraic topology* tries to measure degrees of connectivity using algebraic constructs such as *homology* and *homotopy* groups. *Differential topology* deals with differentiable functions on differentiable manifolds. It is closely related to differential geometry and together they make up the geometric theory of differentiable manifolds. *Geometric topology* primarily studies manifolds and their embeddings in other manifolds. In the context of this thesis, we are primarily interested in algebraic topology, specifically through the concept of homology. We also deal with concepts derived from specific areas of differential topology. These include studying the behaviour of functions on the manifold, and the geometric properties of the manifold as induced by these functions, through Morse theory (Milnor 1963). Specifically, we concentrate on the cosmic density fields, and their geometric and topological properties.

The topological invariant that brings together these different sub-branches of topology is the *Euler characteristic*. It establishes profound, and perhaps even surpris-

ing links between seemingly widely different areas of mathematics (Adler & Taylor 2010). While in simplicial topology Euler's *polyhedron formula* states that the Euler characteristic of a manifold is the alternating sum of the number of k-dimensional simplices needed to triangulate it (Edelsbrunner & Harer 2010)

$$\chi = V - E + F. \tag{1.44}$$

Its role in *algebraic topology* as the alternating sum of Betti numbers is expressed by the Euler-Poincaré formula (see Section 1.6.1 for a definition of the Betti numbers)

$$\chi \;=\; \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{1.45}$$

Even more intricate is the connection that it establishes between these topological aspects and the singularity structure of a field, which is the realm of *differential topology*. In particular, interesting is the relation established by Morse theory of the Euler characteristic being equal to the alternating sum of the number of singularities of different indices, i.e. of maxima, minima and saddle points

$$\chi \;=\; N_{max} - N_{2-saddle} + N_{1-saddle} - N_{min}. \tag{1.46}$$

Finally, its significance in *integral geometry* is elucidated via Crofton's intersection formula (Crofton 1868; Hadwiger 1957), which establishes the fact that Minkowski functionals of a manifold $\mathbb{M}$ can be expressed as integrals over the Euler characteristic of its intersection with hyperplanes in different dimensions. To evaluate the k-th Minkowski functional of a d-dimensional manifold $\mathbb{M}$, one has to consider the Euler characteristic of the intersection of k-dimensional hyperplanes S_k with $\mathbb{M}$, $\chi(S_k \cap M)$. The value of the Minkowski functional $Q_k(\mathbb{M})$ is equal to the integral of the Euler characteristic $\chi(S_k \cap M)$ over the space $\mathcal{E}_k^d$ of all conceivable hyperplanes S_k,

$$Q_k(\mathbb{M}) \;=\; \frac{\omega_d}{\omega_{d-k}\omega_k} \int_{\mathcal{E}_k^d} \mathrm{d}\mu_k(S_k)\, \chi(S_k \cap M)\,, \tag{1.47}$$

where, the normalization constants ω_j are the volumes of j-dimensional unit spheres.

A key aspect of topology that makes it an important tool for studying the cosmic mass distribution is its insensitivity to systematic effects such as compression, stretching, bending, rotation and translation. This property can be immediately used to infer that the topology of the primordial field and that of the evolving field in the later epochs would be invariant and are insensitive to systematic effects such as gravitational evolution, galaxy biasing, and redshift-space distortion (Park & Kim 2010). This is true as long it involves linear and quasi-linear structures, which, formally, is the regime before *shell crossing* occurs.

Topological studies have been invoked to characterize the cosmic mass distribution analysis for many decades. The early studies focused on topological characterization of the cosmic mass distribution by evaluating the genus and the Euler characteristic of the iso-density surfaces, as a function of the density threshold (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986). Later, more discriminative topological information became available with the

introduction of Minkowski functionals (Mecke et al. 1994). As the Euler characteristic is one of the Minkowski functionals, these descriptions are intimately connected. In the last decade, methods emanating from topological formalisms have also provided recipes for detecting and delineating the structures of the cosmic web (Platen et al. 2007; Aragón-Calvo et al. 2010; Sousbie 2011; Sousbie et al. 2011).

In this section, we present a brief description of the topological measures that we mention in the previous paragraphs, specifically the Euler characteristic and the related concept of topological genus. We also discuss their applications in a cosmological context. This establishes the background and context for the introduction of the new topological formalisms described in this thesis. These formalisms are interesting in the sense that they complement and expand on the existing techniques, and as such present a more detailed account of the topology of cosmic mass distribution.

1.5.1 Genus and Euler characteristic

For a connected, orientable surface, the *Genus* [2] is defined as the maximal number of independent simple closed curves that can be drawn on the surface without rendering it disconnected.

We give an intuitive interpretation of genus through the illustration of a torus and a sphere in Figure 1.7. There can be two independent incisions that can be made on the surface of a torus, without rendering it disconnected. These incisions are along the curves drawn in red and blue. This suggests that the genus of a torus is 2. On the other hand, it is not possible to draw any closed curve on the surface of a sphere without rendering it disconnected. Hence, the genus of a sphere is 0. Gott and collaborators (Gott et al. 1986; Hamilton et al. 1986) introduced the use of genus to characterize the cosmic mass distribution.

The Euler characteristic is best understood through the Poincaré formula. Let a closed surface have genus g. Then the *Poincaré formula* states that

$$\chi(\mathbb{M}) = V - E + F, \tag{1.48}$$

is the Euler characteristic, sometimes also known as the Euler-Poincaré characteristic. V, E and F are the number of vertices, edges and faces used to triangulate the manifold. In a more general language, the Euler characteristic of a manifold is the alternating sum of simplex numbers of all the dimensions needed to triangulate the manifold. As an example, let us consider a sphere. Topologically, this is equivalent to a tetrahedron The tetrahedron has 4 vertices, 6 edges, and 4 faces or triangles. So, the Euler characteristic of a sphere (or a tetrahedron) is 2. A vertex, an edge, a triangle and a tetrahedron are examples of a *simplex* in 0-, 1-, 2- and 3-dimensions respectively (see Figure 1.8 for an illustration of simplices in dimensions up to 3), with generalizations existing in higher dimensions.

[2] For consistency, it is important to note that the definition of genus g used in cosmological studies is different from the mathematical definition of genus G, the difference being the number of connected regions: $g = G - c$.

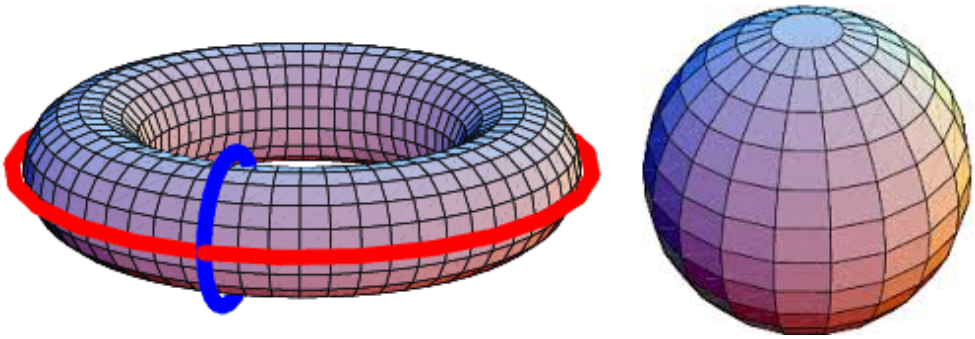


Figure 1.7 Illustration of a torus and a sphere. Genus is the number of independent closed curves that can be drawn on a surface, without rendering it disconnected. For a torus, there can be two independent closed curves that can be drawn without rendering it disconnected (drawn in red and blue). This suggests that the genus of a torus is 2. On the other hand, it is not possible to draw any closed curve on the surface of a sphere without rendering it disconnected. Hence, the genus of a sphere is 0.

Connecting geometry and topology: the Gauss-Bonnet theorem

The Gauss-Bonnet theorem is perhaps the most fundamental theorem that illuminates the connection between geometry and topology. It relates the total Gaussian curvature of a manifold, which is a geometric quantity, to its Euler characteristic or genus, which are topological quantities. The *Gaussian curvature*, κ, of a manifold is given by

$$\kappa = \kappa_1 \kappa_2 = \oint \frac{1}{R_1 R_2} \, dS, \tag{1.49}$$

where, R_1 and R_2 are the radii of curvature of the boundary surface S. κ_1 and κ_2 are the principal curvatures. The Gauss-Bonnet theorem states that (Gauss 1900; Bonnet 1848; Gott et al. 1986)

$$\kappa = 4\pi(1 - g) = 2\pi\chi, \tag{1.50}$$

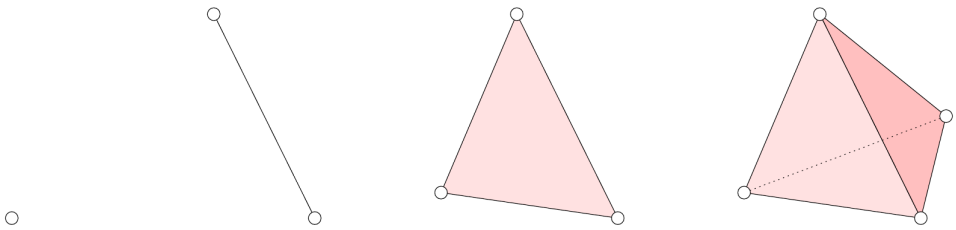


Figure 1.8 From left to right: a vertex, an edge, a triangle, and a tetrahedron. The boundary of a simplex in d dimension is composed of simplices in $d - 1$ dimension. The boundary of an edge has two vertices, the boundary of a triangle has three edges, and the boundary of a tetrahedron has four triangles as faces.

where, κ is the total Gaussian curvature, g is the genus and χ is the Euler characteristic. This suggests that the knowledge of either κ, g or χ is sufficient to compute the others.

The Gauss-Bonnet theorem holds a place of immense importance in the field of computational topology. This is because invoking the Gauss-Bonnet theorem transforms the considerably challenging task of determining the genus or the Euler characteristic, into determining the intrinsic curvature of a manifold. The genus and the Euler characteristic are topological quantities, while the curvature is a geometric quantity. This is a more palatable problem because curvature is easily computable in most situations.

Euler characteristic: Gaussian fields

The genus and the Euler characteristic studies have been an important focal point of topological studies in cosmology. One reason for this is because the analytic closed form expression of the genus and the Euler characteristic, in the case of Gaussian random fields, for iso-density surfaces as a function of density threshold is well known. For Gaussian fields, Gott et al. (1986) and Hamilton et al. (1986) derive the expression for the genus, by computing the average Gaussian curvature, κ, per unit volume, given by

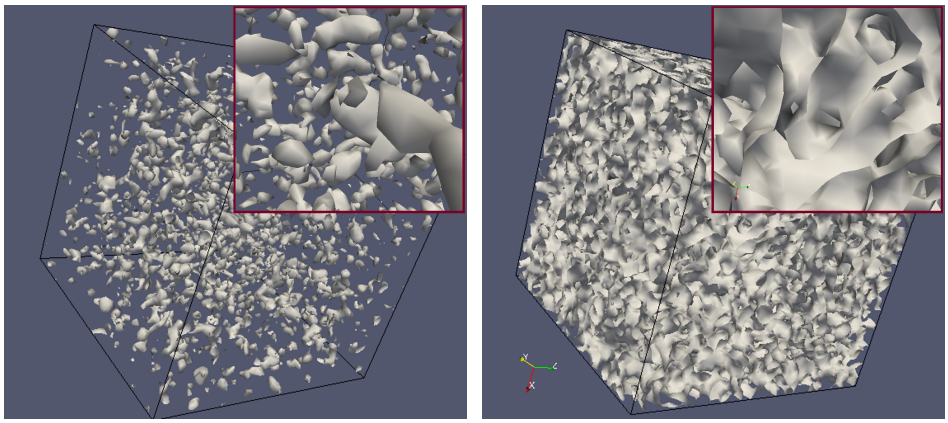
$$\kappa = \frac{1}{\pi}\left[\frac{\ddot{\xi}(0)}{\xi(0)}\right]^{3/2}(1-\nu^2)e^{-\nu^2/2}. \tag{1.51}$$

where, $\xi(0)$ is the auto correlation function, and $\nu = \delta/\sigma$ is the dimensionless density threshold. It denotes the number of standard deviations that the density contrast at a particular location differs from the mean. Thereafter, they relate the calculated expression of the Gaussian curvature to the genus through the Gauss-Bonnet theorem.

The expression for the Euler characteristic has been computed independently by Doroshkevich (1970), Adler (1981) and Bardeen et al. (1986), without invoking the Gauss-Bonnet theorem. This is made possible by recognizing that the Euler characteristic is the alternating sum of the number of critical points of different indices. This very important result follows from a theorem due to Morse (more on Morse theory later) (Milnor 1963). The number (density) of critical points in the case of Gaussian fields is computed independently from appropriate integrals of the probabilistic equation of the Gaussian fields. The expression for the Euler characteristic is given by (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986)

$$\chi(\nu) \;=\; -\frac{1}{8\pi^2}\left(\frac{\langle k^2\rangle}{3}\right)^{3/2}(1-\nu^2)e^{-\nu^2/2}, \tag{1.52}$$

Important to note is that functional form of the genus and the Euler characteristic for Gaussian fields is independent of the specification of the power spectrum, and is a function only of the dimensionless density threshold ν. The contribution from the power spectrum is restricted to the amplitude of the genus curve through the



(a) (b)

Figure 1.9 Figure illustrating *meatball-like* and *sponge-like* topologies. The manifold in panel (a) consists of many isolated components, and hence the analogy with meatballs. The manifold in panel (b) is composed of a single connected surface with many tunnels indented in it. This is much like the topological structure of sponge.

quantity $\langle k^2 \rangle$, which is related to the second moment of the power spectrum, or the auto correlation function. Since the analytic expression for the genus and the Euler characteristic of Gaussian fields is well known, this makes them an ideal tool for testing the hypothesis of initial Gaussian conditions, by comparing the observational data with respect to the analytic formula.

Euler, genus and cosmology

The genus and the Euler characteristic quantifications have also been used routinely to describe the topology of the primordial density field and the large scale structures that emerge out of them at later epochs (Gott et al. 1986; Weinberg et al. 1987; Weinberg 1988; Gott et al. 1989, 1990; Moore et al. 1992; Vogeley et al. 1994; Protogeros & Weinberg 1997; Canavezes et al. 1998; Park et al. 1998, 2001; Hikage et al. 2002; Canavezes & Efstathiou 2004; Park et al. 2005b,a; Gott et al. 2008; James et al. 2009; Gott et al. 2009; Choi et al. 2010).

In addition to testing the initial Gaussian hypothesis, genus topology has also been instructive in illuminating the topological structure of the matter distribution as a function of density threshold. For Gaussian fields, for a high density thresholds, the genus is negative, suggesting a *meat-ball* like topology (Gott et al. 1986; Hamilton et al. 1986). In this case, the manifold is composed of many disjoint pieces and points to a mass distribution with isolated clusters. As an example, the manifold in the panel (a) of Figure 1.9 consists of many isolated components, and hence the analogy with meatballs. For the intermediate thresholds, the genus acquires positive values, and suggests the presence of a *sponge-like* topology (Gott et al. 1986; Hamilton et al. 1986). As an example, the manifold in the panel (b) of Figure 1.9 is composed of a single connected surface with many tunnels indented in it. This is much like

the topological structure of sponge. For low threshold values, the genus again becomes negative. However, the topology is *cheese-like*, pointing to a single connected region with enclosed cavities (Gott et al. 1986; Hamilton et al. 1986). In this case, the topological structure can be imagined as the complement of the manifold depicted in panel (a). The empty regions form a closed connected surface, marked by cavities in the places where the isolated islands are.

1.5.2 Minkowski functionals

There are $(d+1)$ Minkowski functionals, $Q_k((k = 0, \ldots, d)$, for a d-dimensional manifold (Munkres 1984; Mecke et al. 1994; Schmalzing & Buchert 1997; Sahni et al. 1998; Schmalzing et al. 1999; Edelsbrunner & Harer 2010). The first four Minkowski functionals ($d \leq 3$) are the volume, surface area, integrated mean curvature or total contour length, and the Gaussian curvature. Consider a 3-dimensional space $\mathbf{x}$. We calculate the Minkowski functionals for the manifold M_ν, which is the subset of the region $\mathbf{x}$ with density above the threshold ν.

For a given density field, the volume functional $Q_0(\nu)$ is the fractional volume of the regions with density above the threshold ν, normalized by the total volume of the region V. It is given by the volume integral of the Heaviside step function Θ, normalized by the whole volume V (Schmalzing & Buchert 1997)

$$Q_0(\nu) = \frac{1}{V} \int_V \Theta(\nu - \nu(\mathbf{x})). \tag{1.53}$$

The other Minkowski functionals of M_ν can be calculated by appropriate surface integrals of the boundary of M_ν, denoted by $\partial(M_\nu)$. The second Minkowski functional, or the area functional, is given by

$$Q_1(\nu) = \frac{1}{6V} \int_{\partial M_\nu} d^2S(\mathbf{x}). \tag{1.54}$$

The third and the fourth Minkowski functionals, namely the integrated mean curvature functional or the total contour length, and the Gaussian curvature, involve the inverse of the radii of curvatures R_1 and R_2 of the surfaces oriented towards lower density values. The integrated mean curvature functional is given by

$$Q_2(\nu) = \frac{1}{6\pi V} \int_{\partial M_\nu} d^2S(\mathbf{x})[\kappa_1 + \kappa_2], \tag{1.55}$$

and the Euler characteristic is given by

$$Q_3(\nu) = \frac{1}{4\pi V} \int_{\partial M_\nu} d^2S(\mathbf{x})[\kappa_1\, \kappa_2], \tag{1.56}$$

where $\kappa_1 = 1/R_1$ and $\kappa_2 = 1/R_2$.

Minkowski functionals are predominantly geometric in nature, though there are

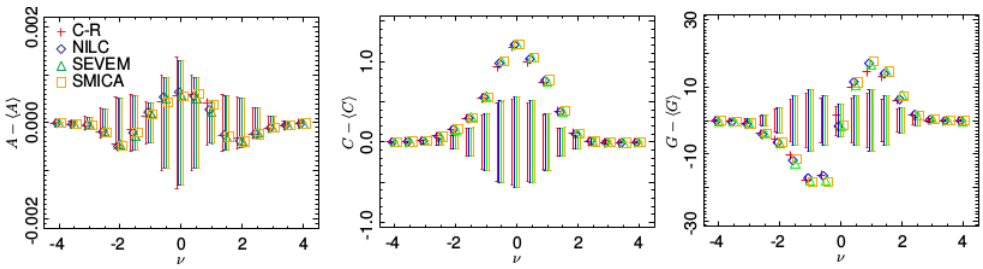


Figure 1.10 Difference between the Minkowski Functionals of the CMB (unnormalized), as measured by PLANCK, with respect to the average of the curves obtained with realistic simulations of the CMB for several cleaned maps. From left to right: Area, Contour, Genus. The error-bars represent the 1σ (68%CL) dispersions around the mean obtained with simulations. Figure courtesy Planck Collaboration et al. (2013a).

connections to topology as well, through the Euler characteristic. The first three Minkowski functionals describe the geometry of the manifold. The volume functional $Q_0(\nu)$ computes the fractional volume above the density threshold ν. The surface area functional $Q_1(\nu)$ and integrated mean curvature functional $Q_2(\nu)$ can be used to characterize the morphology of the manifold. Pancake or wall-like features are characterized by a large surface area and small integrated mean curvature. On the other hand, filamentary regions are characterized by small surface area and a large integrated mean curvature.

Minkowski functionals were introduced as measures of the spatial cosmic mass distribution by Mecke et al. (1994) and have become an important measure of clustering of mass and galaxies (Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998).

In cosmology, Minkowski functionals have primarily been used for model discrimination. In this context, they have been an important focal point of topological and morphological studies, because their analytic closed form expression in the case of Gaussian random fields for iso-density surfaces as a function of density threshold is well known (Tomita 1993; Schmalzing & Buchert 1997). This makes them an ideal tool for testing the hypothesis of initial Gaussian conditions through comparison with observational data. For Gaussian random fields, the expected value of the first four Minkowski functionals of the excursion sets have known analytical expressions (Tomita 1993; Schmalzing & Buchert 1997)

$$
\begin{aligned}
Q_0(\nu) &= \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{1}{\sqrt{2}}\nu\right), \\
Q_1(\nu) &= \frac{2}{3}\frac{\lambda}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\nu^2\right), \\
Q_2(\nu) &= \frac{2}{3}\frac{\lambda^2}{\sqrt{2\pi}}\nu\exp\left(-\frac{1}{2}\nu^2\right), \\
Q_3(\nu) &= \frac{\lambda^3}{\sqrt{2\pi}}(\nu^2-1)\exp\left(-\frac{1}{2}\nu^2\right).
\end{aligned} \tag{1.57}
$$

where $\lambda = \sqrt{|\xi''(0)|/[2\pi\xi(0)]}$ is related to the correlation function or the power spectrum, and the function $\Phi(x) = \int_0^x \mathrm{d}te^{(-t^2)}$ is the standard error function.

The Minkowski functionals of the CMB have also been calculated and compared with the Minkowski functionals of theoretical Gaussian random field curves to test the Gaussian initial condition hypothesis. Figure 1.10 presents the difference of the data Minkowski Functionals from the Planck measurement of the CMB (unnormalized) with respect to the average of the curves obtained with realistic Planck simulations for several cleaned maps (Planck Collaboration et al. 2013a). From left to right are plotted the Area, Contour and Genus. The error-bars represent the 1σ (68%CL) dispersions around the mean obtained with simulations. From this figure, a deviation at a level of $\sim 2\sigma$ can be seen for the contour and genus curves. The authors claim that the difference is not compelling, and that the back- ground of unresolved sources may be responsible for at least part of the excess signal that is detected.

Minkowski functionals as shape finders. That the Minkowski functionals have predominantly been recognized as related to the geometric properties of a manifold can also be appreciated from the fact that specific combinations of Minkowski functionals, called *shapefinders* (Sahni et al. 1998) can be used to provide a set of non-parametric measures of sizes and shapes of objects. These are the $\mathscr{H}_1$ (Thickness), $\mathscr{H}_2$ (width) and $\mathscr{H}_3$ (Length), defined as follows (Sahni et al. 1998):

$$
\mathscr{H}_1 = \frac{V}{A}, \mathscr{H}_2 = \frac{A}{C}, \mathscr{H}_3 = C. \tag{1.58}
$$

The above three shapefinders can also be used to construct a pair of another dimensionless shapefinders, given by (Sahni et al. 1998)l

$$
\mathscr{K}_1 \equiv \frac{\mathscr{H}_2 - \mathscr{H}_1}{\mathscr{H}_2 + \mathscr{H}_1}; \quad \mathscr{K}_2 \equiv \frac{\mathscr{H}_3 - \mathscr{H}_2}{\mathscr{H}_3 + \mathscr{H}_2} \tag{1.59}
$$

The set of shapefinders ($\mathscr{H}_1$,$\mathscr{H}_2$,$\mathscr{H}_3$), as well as ($\mathscr{K}_1$,$\mathscr{K}_2$) describe the shape of a given region. ($\mathscr{H}_1$,$\mathscr{H}_2$,$\mathscr{H}_3$) have dimensions of length and provide an estimate of the "extension" of the region: $\mathscr{H}_1$ is the shortest and thus describes the characteristic thickness of the region or object; $\mathscr{H}_2$ is intermediate and can be associated with

the breadth of the object; $\mathscr{H}_3$ is typically the longest and characterizes the length of the object. $\boldsymbol{\mathscr{H}} = (\mathscr{H}_1, \mathscr{H}_2)$ can be regarded as a two-dimensional vector whose amplitude and direction determine the shape of an arbitrary three-dimensional surface (Sahni et al. 1998). An ideal pancake, which may have vanishing thickness but is not necessarily planar, has one characteristic dimension much smaller than the remaining two, so that $\mathscr{H}_1 \ll \mathscr{H}_2 \simeq \mathscr{H}_3$ and $\boldsymbol{\mathscr{H}} \simeq (1, 0)$. A typical filament, which is a one-dimensional object but not necessarily straight, has two characteristic dimensions much smaller than the third, so that $\mathscr{H}_1 \simeq \mathscr{H}_2 \ll \mathscr{H}_3$ and $\boldsymbol{\mathscr{H}} \simeq (0, 1)$. All three dimensions of a sphere are equal, resulting in $\mathscr{H}_1 \simeq \mathscr{H}_2 \simeq \mathscr{H}_3$ and $\boldsymbol{\mathscr{H}} \simeq (0, 0)$. Important to note is that, the shapefinders only describe the morphology of simply connected shapes. Sahni et al. (1998) note that for multiply connected regions, the knowledge of genus is also essential to fully characterize it. The dimensionless triad ($\mathscr{K}_1$,$\mathscr{K}_2$, g), gives information about shape as well as topology.

1.6 Beyond the Euler characteristic and Minkowski functionals

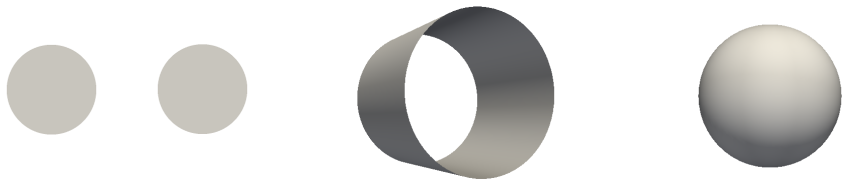
Genus and Euler characteristic, along with the rest of the Minkowski functionals, have provided crucial insights into the topology and morphology of cosmic density fields from models as well as observational data. For example, the Euler characteristic of the cosmic density fields in the linear regime suggest the presence of near-Gaussian conditions in the early Universe. Nonetheless, there are a range of grounds that make it necessary to explore a more elaborate and extensive arsenal of topological concepts and descriptions.

- While the genus can distinguish between connected, closed surfaces in $\mathbb{R}^3$, it has no discriminative power if applied to a 3-manifold. In other words, the genus is defined only for connected and closed 2-dimensional surfaces, and has no generalizations in higher or lower dimensions.

- The Euler-Poincaré formula states that the Euler characteristic is an alternating sum of another topological invariant called the *Betti numbers* (Betti 1871). While, we will elaborate on the Betti numbers in a later section, it suffices to say here that the Euler characteristic represents a more compressed form of topological information, than what the Betti numbers may supply.

- It is established that structures in the Universe form and evolve in a hierarchical fashion. Smaller high density structures coalesce together hierarchically to build up larger structures of lower density. A topological description through genus and Euler characteristic is not equipped to address this aspect of the cosmic mass distribution. A formalism capable of expressing topology in a hierarchical fashion would present an interesting and powerful alternative for describing the hierarchical structures in the cosmos.

In view of the above observations, this thesis describes the topology of the models of cosmic mass distribution through homology, Morse theory and persistence. These formalisms, as well as techniques derived from them, capture the topology of



(a) (b) (c)

Figure 1.11 Figure illustrating the concept of topological holes. (a) There are two independent isolated discs, which are examples of 0-cycles.The gap thats formed due to these discs (all the space, except the discs) is a 0-dimensional hole. (b) The cylinder is an example of a 1-cycle (loop) that encloses a tunnel. (c) The surface of a (empty)-sphere is an example of a 2-cycle, that encloses a void. In addition, (b) and (c) also exhibit the presence of a 0-cycle, because the circle and the surface of a sphere are both connected objects.

a manifold in a more detailed fashion than Euler characteristic and genus. Additionally, persistence has the ability to describe topological information in a hierarchical fashion.

The topological structure of a manifold can also be understood via the path of differential topology. The theoretical framework that deals with this aspect is called Morse theory (Milnor 1963). Morse theory is the study of the singularity structure of a manifold as induced by a smooth scalar function defined on it. The nature of the *critical points* of the function, i.e. locations where the gradient of the function vanishes, decides the kind of topological holes that may form in the manifold. Due to its intimate connection with topological holes, Morse theory is also directly relevant towards developing an intuitive understanding of persistence. Another crucial aspect of Morse theory is that the geometric properties of the manifold in the neighborhood of the critical points can be harnessed to design structure detection algorithms, that can be used to delineate the structures in the cosmic web.

We describe the fundamentals of homology, Morse theory and persistence in the next subsections.

1.6.1 Homology and Betti Numbers

Homology is a mathematical formalism for specifying in a quantitative and unambiguous manner about how a space is connected [3], through assessing the boundaries of a manifold (Munkres 1984). Homology groups provide a mathematical language for the holes in a topological space (Edelsbrunner & Harer 2010). A d-manifold can be composed of topological holes of 0 up to $(d-1)$ dimensions. Holes in $d < 3$ have intuitive interpretations. A 0-dimensional hole is a *gap* between two isolated independent objects. A 1-dimensional hole is a *tunnel* through which one can pass in any one direction without encountering a boundary. A 2-dimensional hole is a *cavity* or

[3]There is a notion of k-connectedness, $k = 0, \ldots, d$, where d is the dimension of the manifold. Within this, 0-connectedness is the same as the 'usual' notion of connectedness.

void fully enclosed within a 2-dimensional surface. This intuitive interpretation in terms of 'gaps' and 'tunnels' is only valid for surfaces embedded in $\mathbb{R}^3$, $\mathbb{S}^3$ or $\mathbb{T}^3$.

Alternatively, one can also talk about these holes in terms of what surrounds them. This is through defining the holes via *cycles*. A 0-cycle is a connected object (and hence, a 0-hole is the gap between two independent objects). A 1-cycle is a *loop* that surrounds a tunnel. A 2-cycle is a *shell* enclosing a void.

The collection of all p-dimensional cycles is the p-th homology group $\mathbb{H}_p$ of the manifold. The rank of this group is the collection of all *independent* cycles. The rank is denoted by the *Betti numbers* β_p, where $p = 0, \ldots, d$ (Betti 1871; Edelsbrunner & Harer 2010). The first three Betti numbers have intuitive meanings. β_0 counts the number of independent components, β_1 counts the number of loops enclosing the independent tunnels and β_2 counts the number of shells enclosing the independent voids.

In panel (a) of Figure 1.11, $(\beta_0, \beta_1, \beta_2) = (2, 0, 0)$, because there are two isolated objects, and no additional holes in in any other dimension. In panel (b), $(\beta_0, \beta_1, \beta_2) = (1, 1, 0)$, because there is a single connected object (circle), which also bounds a tunnel. In panel (c), $(\beta_0, \beta_1, \beta_2) = (1, 0, 1)$, because the surface of a sphere is a connected object, which also bounds a void. A more mathematically rigorous definition of homology groups and Betti numbers can be found in Appendix A.3 as well as more traditional literature on the subject, like e.g. by Munkres (1984).

Figure 1.11 illustrates the concept of topological holes. In panel (a), there are two independent isolated discs, which are examples of 0-cycles.The gap thats formed due to these discs (all the space, except the discs) is a 0-dimensional hole. In panel (b), the tunnel is an example of a 1-cycle (loop) that encloses a tunnel. In panel (c), the surface of a (empty)-sphere is an example of a 2-cycle, that encloses a void. In addition, panels (b) and (c) also exhibit the presence of a 0-cycle, because the circle and the surface of a sphere are both connected objects.

Like the Euler characteristic, the Betti numbers are topological invariants of a manifold, meaning that they do not change under systematic transformations like rotation, translation and deformation. Betti numbers contain strictly more topological information than the Euler characteristic. This is because the Euler characteristic of a manifold can be written as an alternating sum of the Betti numbers. The relationship to the Euler characteristic is given by the *Euler-Poincaré Formula* (Edelsbrunner & Harer 2010; Adler & Taylor 2010):

$$\chi \;=\; \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{1.60}$$

where d is the intrinsic dimension of the manifold.

This information is of key importance because it tells us immediately that it is possible for two different manifolds to have the same Euler characteristic but different sets of Betti numbers. As a result manifolds which are topologically equivalent through a description of Euler characteristic may turn out to have different topologies when described through the Betti numbers. This has important repercussions for a topological description of the cosmic mass distribution when model discrimination is the primary focus. A crucial point to note in this context is that Wintraecken & Vegter (2013) show that Betti numbers cannot be formulated as integrals of the Euler

characteristic.

1.6.2 Morse theory

Morse theory is the study of the topology of the level sets of a manifold via the critical points of a smooth scalar function defined on it. Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be a real-valued scalar function defined on a manifold $\mathbb{M}$. *Critical points* of f are points of f where the gradient of f vanishes i.e., $\nabla f = 0$. The function f is said to be a *Morse function* if all of its critical points are non-degenerate i.e., the *Hessian* of f, defined as the matrix of second order partial derivatives, is non-singular. The non-degeneracy condition imposes a locally quadratic form for f, within a small neighborhood of its critical points (Milnor 1963). In other words, using a coordinate transformation, the function near a critical point p of the d-dimensional manifold $\mathbb{M}$ can be written as a quadratic function

$$f_p(x) = f(p) \pm x_1^2 \pm x_2^2 \pm ... \pm x_d^2. \tag{1.61}$$

The *index* of p is equal to the number of negative quadratic terms in the above expression. In 2D, the index 0 corresponds to a minimum, the index 1 corresponds to a saddle point and index 2 corresponds to a maximum. The saddle point can be of two kind, namely the normal saddle and the monkey saddle. In 3D, the index 0 corresponds to minima, the index 1 corresponds to 1-saddles, the index 2 corresponds to 2-saddles, and the index 3 corresponds to maxima. Figures 1.12 and 1.13 illustrate the geometry as the behavior of the gradients of the manifolds in the neighborhood of various kinds of critical points in 2D and 3D.

Critical points and topology

A major result that Morse theory establishes is that the global shape and topology of a manifold is defined only by it critical points (Milnor 1963; Edelsbrunner & Harer 2010). To put it more formally, let $\mathbb{M}$ be a closed manifold and $f : \mathbb{M} \rightarrow \mathbb{R}$ be a Morse function on M. Let $\mathbb{M}_t = \{p \in \mathbb{M} | f(p) \leq t\}$ for any arbitrary t in f. If f has no critical points in $[a, b] \in \mathbb{R}$, then $\mathbb{M}_a$ and $\mathbb{M}_b$ are diffeomorphic, or topologically equivalent. This is a very important result, and implies that all changes in topology of $\mathbb{M}$ occur only at the critical points of f. Theoretically, there are infinitely many levels of density as one decreases the function value from the highest to the lowest. But fortunately, the topology of the manifold only changes while passing a critical point. This is of key importance because this means that the infinite number of levels of density threshold can be constrained to a finite number, by only having to consider one level in between any two critical points.

Since there are finitely many critical points c_i, one may order them according to decreasing function values. Then passing from the highest-value critical point to the lowest-value critical point gives all the unique topological changes in the manifold.

For illustration, let us consider the vertical torus, with the height function defined, illustrated in Figure 1.14. Sweeping from the lowest to the highest function values, one passes the critical points of the height function. These are responsible for the topological changes. From left to right: passing the minimum creates a disk,

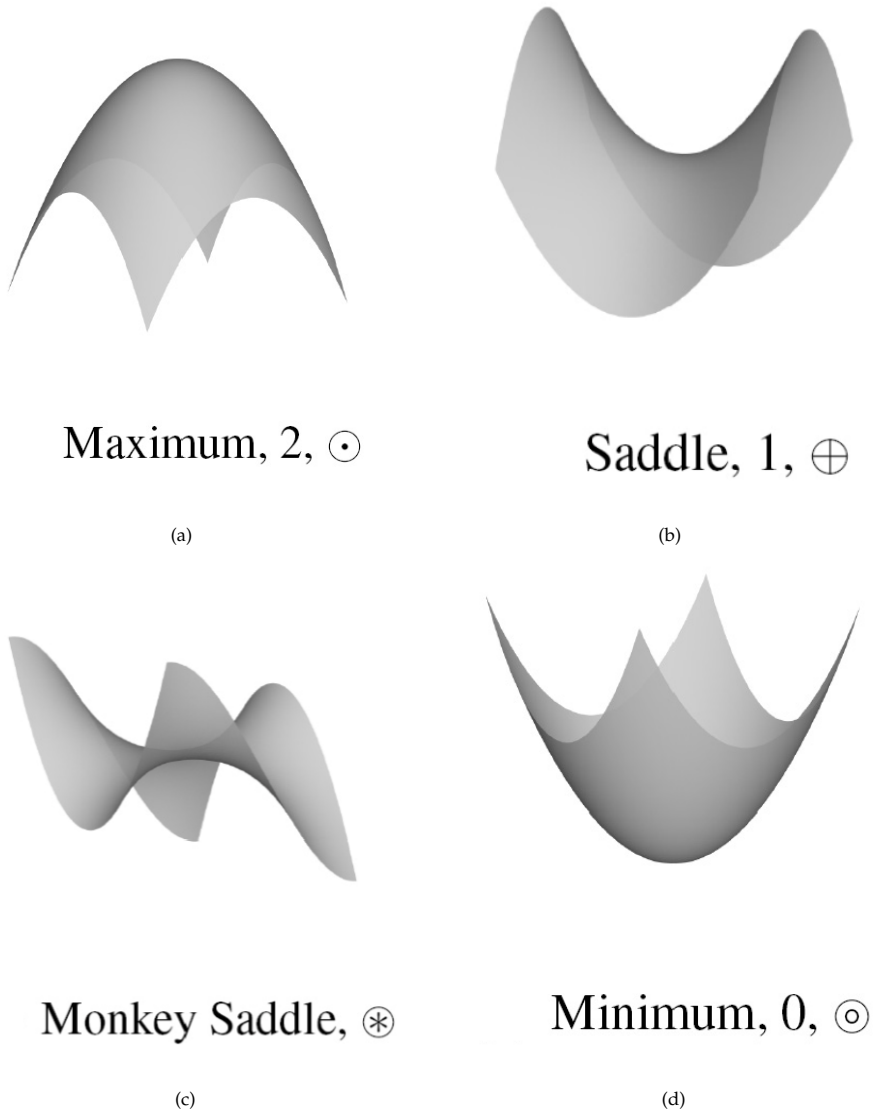


Figure 1.12 Critical points in 2D, and shape of a manifold in the neighborhood of critical points. Index 0 is a minimum, index 1 is a saddle point and index 2 is a maximum.

passing a 1-saddle creates a cylinder, passing another 1-saddle creates a torus with a hole, and passing the maximum caps the hole and generates the whole torus.

A crucial point to note is that the critical points of a function are not only responsible for the *formation* of a topological hole, but also for their *destruction*. We will elaborate on this in the next subsection, and harness these concepts to build an

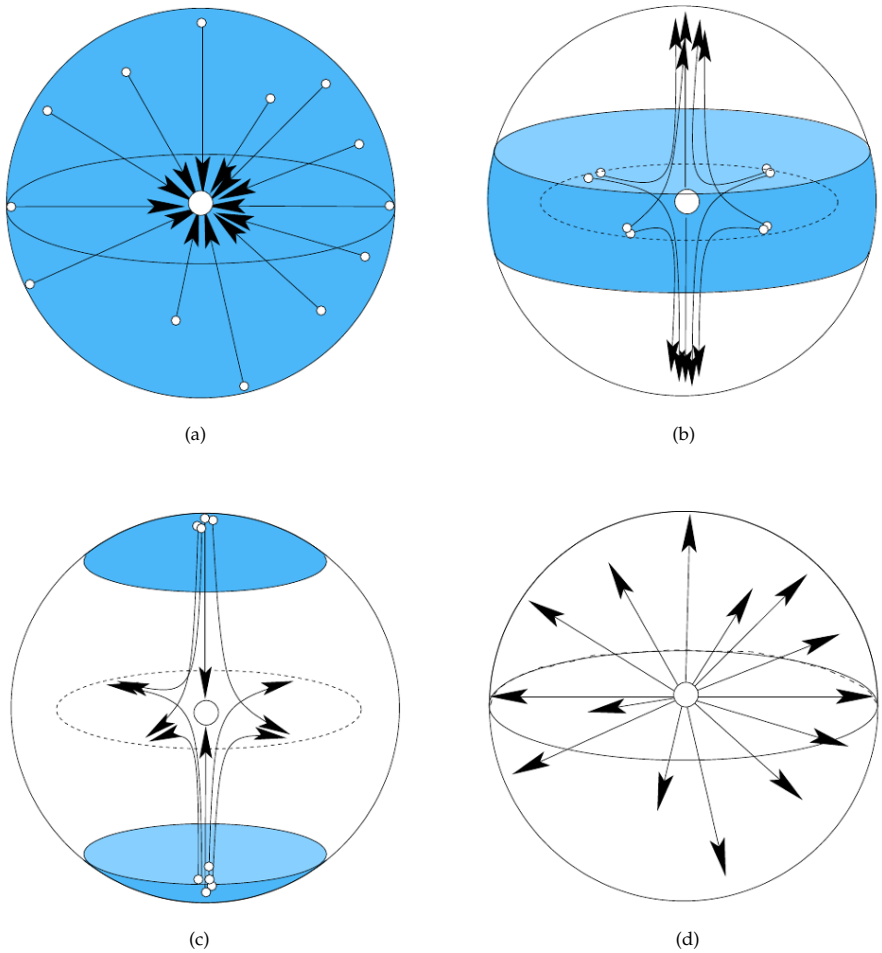


Figure 1.13 Critical points in 3D. Panels (a) through (d) present the maximum, 2-saddle, 1-saddle and the minimum. For a maximum, the flow is directed towards it. For a minimum, the flow is directed away from it. For a 1-saddle, the flow is towards it in a plane, and away from it along a linear element. The directions reverse for a 2-saddle.

intuitive understanding of persistence.

1.6.3 Persistence homology

In Section 1.6.2, we noted the relation between the critical points of a function with the topological changes it induces in a manifold. In this section, we use the notions described above to sketch an intuitive understanding of persistence homology (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Carlsson et al. 2005; Carlsson & Zomorodian 2009; Carlsson 2009; Edelsbrunner & Harer 2010).

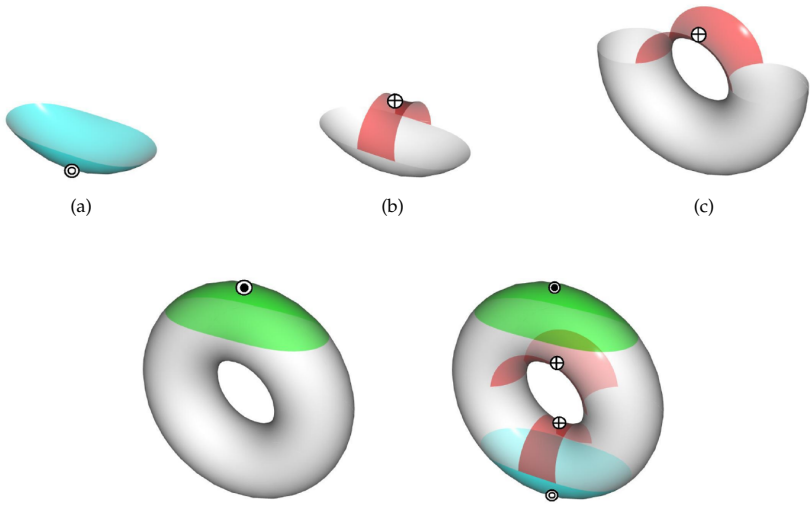


Figure 1.14 A vertical torus with a height function defined. Sweeping from the lowest to the highest function values, one passes the critical points of the height function. These are responsible for the topological changes. From panel (a) to panel (e): passing the minimum creates a disk, passing a 1-saddle creates a cylinder, passing another 1-saddle creates a torus with a hole, and passing the maximum caps the hole and generates the whole torus. In other words, the various kinds of critical points generate the cycles of different dimension of the torus. The minimum generates the 0-cycle (disk). The two 1-saddles generate the two 1-cycles (loops or tunnel), while the maximum generates the 2-cycle. This is the whole torus that encloses the cylindrical void (also one of the tunnels). Figure courtesy Gyulassy (2008).

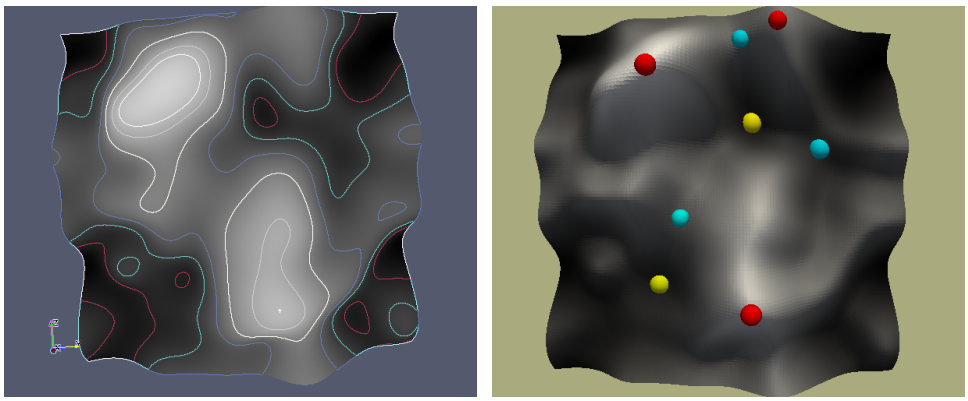


Figure 1.15 (a) Face on view of a 2D random field. A range of level sets are presented as contours of different colors. (b) Surface view of the same. Maxima are marked with red balls, saddles with cyan balls, and minima with yellow balls.

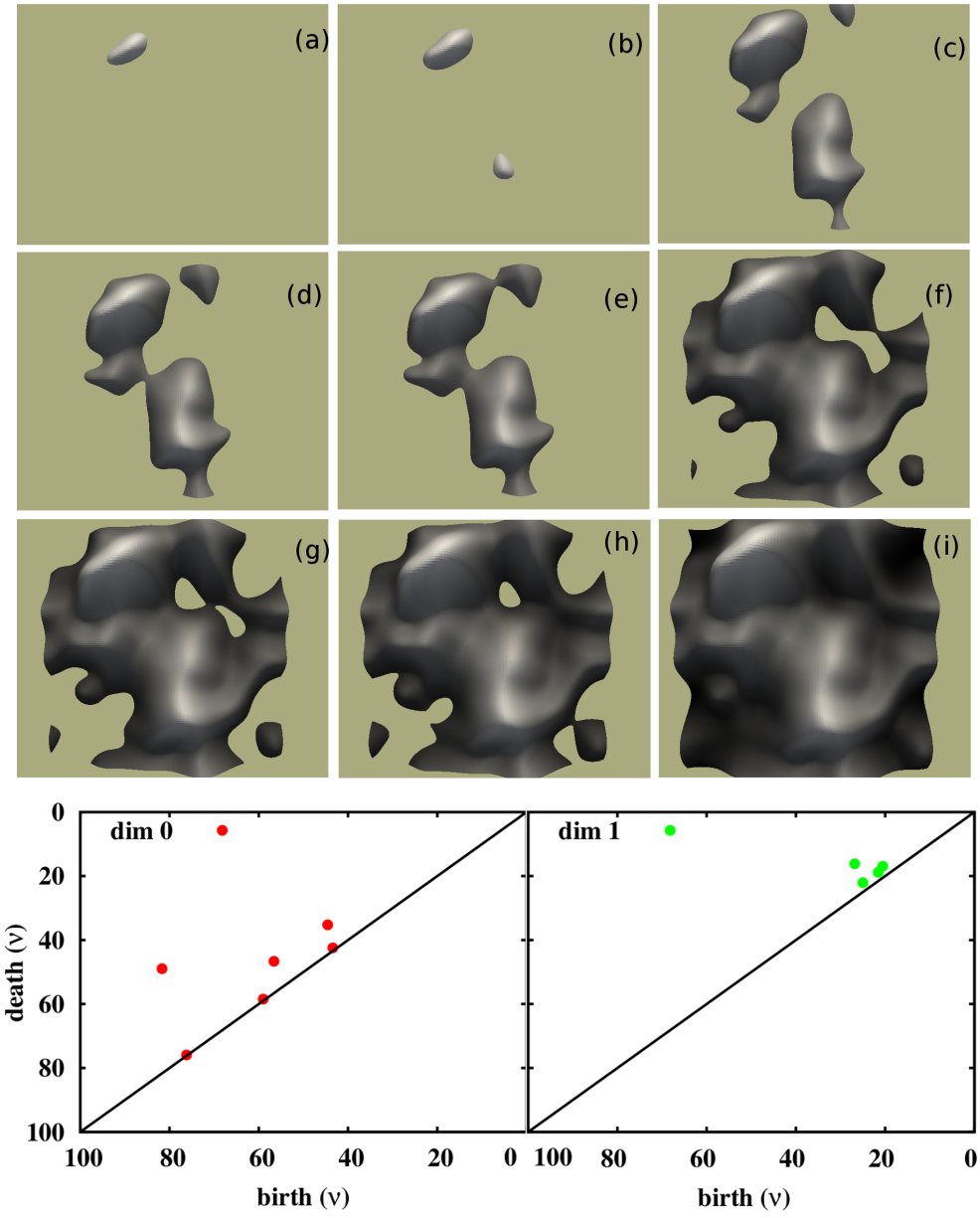


Figure 1.16 Persistence and field singularity structure. The process of birth and death as we grow the superlevel sets by decreasing the density threshold in a given 2-dimensional random field. The events of birth and death are quantified in the persistence diagrams in the bottom row. Bottom-left: 0-dimensional diagram, bottom-right: 1-dimensional diagram.

The key observation is that the topology of the manifold changes only when passing through a critical point of f. Introduction of a critical point to a manifold results in unique topological changes. These changes may result in either the *birth* (formation) or *death* (destruction) of a topological hole. More specifically, the addition of a p-critical point can result in either the birth of a p-dimensional hole or the death of a $(p-1)$-dimensional hole (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Edelsbrunner & Harer 2010). For example, consider a saddle point in a 2D manifold. The introduction of this saddle point can result in either of the following topological changes: it can either connect two disjoint objects, which results in the destruction of a 0-hole; or it can connect the boundary of two already connected objects, forming a loop. This depicts the creation of a 1-dimensional hole or tunnel.

Central to the formulation of persistence is the necessity to track the birth and death events, as one changes the superlevel or sublevel sets. In this thesis, we track the topological changes through changing superlevel sets, noting that tracking it through changing sublevel sets is equivalent.

Figure 1.15 presents a 2D random field, that we use as an example to illustrate the working of persistence. In panel (a), we present the face on view of the field. A range of level sets are presented as contours of different colors. In panel (b), we present the surface view of the same. Maxima are marked with red balls, saddles with cyan balls, and minima with yellow balls. Figure 1.16, illustrates the working of persistence by tracking the birth and death of islands and tunnels, for growing superlevel sets, for the $2D$ random field show in Figure 1.15. We employ a $2D$ field for ease of visualization.

We trace the growing superlevel set from the top-left panel to the bottom right panel, as a function of decreasing thresholds of f. We only show regions of the manifold that are included in the superlevel set. We trace the change in topology of this region, as the superlevel set grows. In panel (a), we start with a single island, which is also a peak. In panels (b) and (c), we witness the birth of two more islands, when two additional peaks get included in the excursion region. In panel (d), two of the islands merge and we are left with two islands as a result. This merger results in the death of one of the islands. This is the island which was born at a later threshold. This is according to the *elder rule* (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Edelsbrunner & Harer 2010), which dictates that given a choice between killing two components, the one that is born later is killed preferentially. In panel (e), there is another merger of two isolated islands.

In panel (f), the first 1-dimensional hole or a loop is born. It has the appearance of a lake surrounded by land. In panel (g), this loop splits into two. In panel (h), one of the loops is filled, while the other one still exists. This indicates the death of a 1-dimensional hole. The density value at which a hole is born, and the density value at which it gets destroyed quantify its *life-span*. The life-span is equal to the difference between the absolute values of the density of birth and death. This means that one of the two loops has a higher life span or persistence than the other. It is likely to be a more significant feature than the other. In panel (i), all the holes fill up, and the superlevel set consists of the whole manifold.

One can record these topological changes as one sweeps from the highest to the

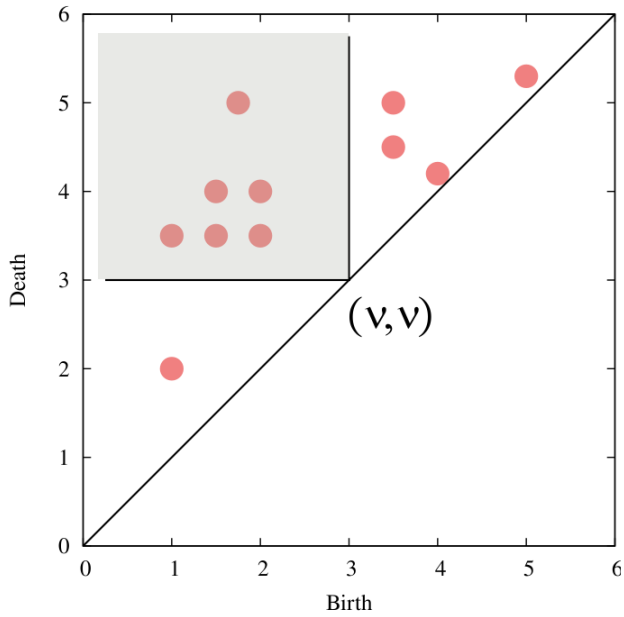


Figure 1.17 The Betti numbers can be read off from the persistence diagrams. The contribution to the Betti numbers for a level set ν comes from all the persistent dots that are born before ν and die after ν – in other words, the shaded region anchored at (ν, ν).

lowest density values, or a continuously decreasing level-set. Each topological hole is associated with two unique function values: $f(c_b)$ associated with the critical point c_b that gives birth to the hole, and $f(c_d)$ associated with the critical point c_d that is responsible for *killing* or filling up the hole. The *life-span* of the hole, or its *persistence*, π, is then given by the absolute difference between the death and the birth values associated with the hole (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Edelsbrunner & Harer 2010)

$$\pi = |f(c_b) - f(c_d)|. \tag{1.62}$$

Persistence diagrams

Persistence homology is represented in terms of *persistence diagrams* (Edelsbrunner & Harer 2010), which is a collection of dots, each dot associated with a unique topological change in the manifold (see for example the bottom row of Figure 1.16). There is a diagram for each ambient dimension of the manifold. 0-dimensional diagrams record the merger events of two isolated objects. 1-dimensional diagrams record the formation and destruction of loops, while 2-dimensional diagrams record the birth and death of topological voids.

Alternatively, the information on persistence topology can also be depicted using *persistence barcodes* (Zomorodian & Carlsson 2005; Carlsson et al. 2005; Carlsson 2009;

Adler et al. 2010). The barcode diagram is a collection of line segments on the real line. Each bar corresponds to a topological feature, like the dots in the dot diagrams. The coordinates of the end points of the bars in the barcode diagrams denote the birth and the death values of the topological features. We note that the dot diagrams and the barcode diagrams are essentially two different representations of the same information.

Persistence diagrams contain more information than the Betti numbers. We can read the p-th Betti number of the superlevel set for ν from the persistence diagram. The contribution to the Betti numbers for a level set ν comes from all the dots in the persistence diagram corresponding to cycles that are born before ν and die after ν – in other words, the shaded region anchored at (ν, ν) in Figure 1.17. Another useful property is the stability of the diagram under small perturbations of the input function. This was pointed out by Cohen-Steiner et al. (2007). Specifically, the diagram of a density function, ϱ', that differs from ϱ by at most ε at every point of the space, has a *bottleneck distance* at most ε from $\mathrm{Dgm}_p(\varrho)$ (Cohen-Steiner et al. 2007). This implies that every points of $\mathrm{Dgm}_p(\varrho')$ is at a distance at most ε from a point in $\mathrm{Dgm}_p(\varrho)$, or from the horizontal axis. The stability result of the persistence diagrams has been used to devise bootstrapping methods for determining the confidence intervals for persistence diagrams arising from stochastic processes (Chazal et al. 2013).

1.6.4 Information hierarchy between Persistence, Homology and Euler characteristic

An important point to note is that persistence, homology and Euler characteristic contain strictly decreasing amount of topological information about the manifold. They are also closely related in the sense that persistence homology contains information about the homology of the manifold, and homology contains information about the Euler characteristic of the manifold. It is not possible to reverse the sequence.

Since persistence homology contains more topological information about the manifold, and is hierarchical in nature, the ideas stemming from it take central importance in the topological characterization of cosmic mass distribution models in this thesis. Together with the Minkowski functionals, persistence homology and homology characterize the geometry and topology of the manifold more completely. As such, one intention of this thesis is to propose them as crucial and essential tools for studying the topology of the cosmic mass distribution.

1.6.5 Feature detection through topology: Segmentation via Morse-Smale complex

Morse theory and related formalisms can be harnessed to devise methods of partitioning a manifold (Forman 2002; Edelsbrunner et al. 2001, 2003; Gyulassy 2008; Gyulassy et al. 2008; Aragon-Calvo et al. 2010; Sousbie 2011). This particularly depends on the geometry of the space in the neighborhood of the critical points of f. Such a partitioning of the domain is achieved by exploring the features of the gradient of the scalar function on the manifold.

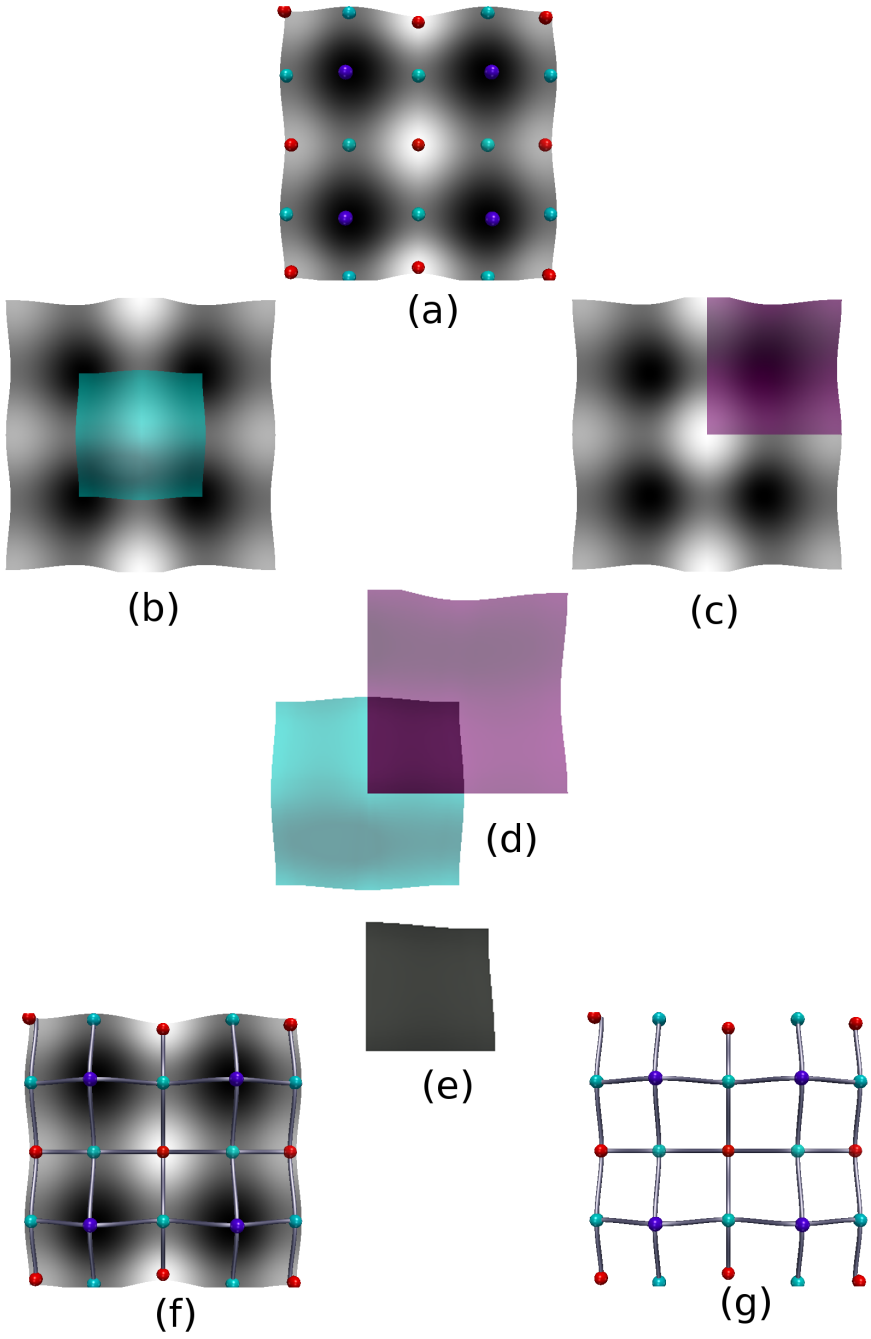


Figure 1.18 Caption next page

Figure 1.18 Decomposition of a terrain into ascending and descending manifolds, and the Morse-Smale complex of the same. (a) The function f along with its singularities. Maxima are depicted by red balls, saddles are depicted by green balls, and minima are depicted by blue balls. (b) Descending manifold of a maximum of f. (c) Ascending manifold of a minimum of f. (d) Intersection of the ascending manifold of the maximum and descending manifold of the minimum. The Morse-Smale complex is the decomposition of a manifold into elements such that the intersection of any two d-dimensional manifolds is a $(d-1)$-dimensional manifold. (e) A 2-dimensional element of the Morse-smale complex. (f) The Morse-Smale complex of the manifold defined by f. The decomposition is illustrated with tubular elements. (g) It is sufficient to have the information about nodes, or the critical points, and the arcs that connect them, in order to be able to fully define a Morse-Smale complex. This is the combinatorial representation of the Morse-Smale complex of f.

An important ingredient in understanding how this partitioning may be achieved is the concept of *integral lines*. Let us consider a manifold $\mathbb{M}$. A curve $l(t)$ is an integral line of f on $\mathbb{M}$, if $\partial l(t) = \nabla f(l(t))$ for all $t \in \mathbb{R}$. In other words, an integral line is a path where the tangent of the path is parallel to the gradient ∇f at every point along the path. Integral lines represent the flow along the gradient between critical points. For example, the lines used to represent the direction of flow in the Figure 1.13 are the integral lines of the domain. The origin and destination of an integral line are the critical points of f. An important property of integral lines is that they cover all of $\mathbb{M}$.

The integral lines can be used to define the *ascending* and *descending* manifolds of a critical point. The set of all integral lines that originate at the critical point p together with p is called the *ascending manifold* of p. Similarly, the set of all integral lines that terminate at the critical point p together with p is called the *descending manifold* of p. The ascending manifolds (similarly, the descending manifolds) of all critical points partition the domain. The ascending manifold of a critical point with index i has dimension $n - i$, where n is the dimension of the domain. Thus, the ascending manifold of a minimum is a three dimensional cell, the ascending manifold of a 1-saddle is a two dimensional sheet, the ascending manifold of a 2-saddle is a one dimensional arc, and the ascending manifold of a maximum is equal to the maximum. The converse is true for the descending manifold i.e., the descending manifold of a critical point with index i has dimension i. Figure 1.13 also presents an illustration of the ascending and descending manifolds associated with the critical points of different indices in 3D. The ascending manifolds of a minimum, 1-saddle, 2-saddle, and maximum are a volume, a surface, pair of lines, and a point respectively. Similarly, the descending manifolds are a point, a pair of lines, a surface, and a volume.

The decomposition of a domain into its ascending or descending manifolds forms the *Morse complex* of f (Forman 2002; Edelsbrunner et al. 2001, 2003). Figure 1.18 presents an illustration of the partitioning of a domain into its Morse-Smale complex. Panel (a) in the figure presents the domain of interest of a function f. It is a 2D surface, where we mark the positions of maxima, minima and saddle points with red, blue and green translucent spheres. Panels (b) and (c) of the Figure 1.18 presents an illustration of an ascending and a descending manifold. The pink surface in panel (b) illustrates a typical descending manifold of a maximum of f. Every point in this region is part of an integral line whose destination is the maximum in the center of

the region. The blue surface in panel (c) illustrates a typical ascending manifold of a minimum of f. Every point in this region is part of an integral line whose origin is the minimum in the center of the region.

Constructing the *Morse-Smale complex* requires the decomposition of a terrain into cells formed by the collection of integral lines that share a common source and a common destination (Forman 2002; Edelsbrunner et al. 2001, 2003). This can only happen if the ascending and descending manifolds of all pairs of critical points intersect only transversally. This means that if the ascending and descending manifolds of two critical points intersect, then the intersection has dimension exactly equal to the difference in the indices of the two critical points. A function f satisfying this condition is called a *Morse-Smale function*. Panel (d) shows the intersection of the ascending and descending manifolds of the minimum and the maximum. Panel (e), shows a 2-cell resulting from this intersection. This is a typical 2-cell of the Morse-smale complex. Panel (f) shows the Morse-smale complex of f. It is sufficient to have the information about nodes, or the critical points, and the arcs that connect them, in order to be able to fully define a Morse-Smale complex. This is the combinatorial representation of the Morse-Smale complex of f. Panel (g) shows the structure of the combinatorial representation of the Morse-Smale complex of f.

The various elements of the cosmic web have natural geometric associations with the elements of the Morse-Smale complex constructed from cosmic density fields. Clusters can be seen as the local maxima of the field, filaments can be identified as the ascending manifold of 2-saddles, voids can be identified as the ascending manifolds of minima, and walls can be identified as the descending manifolds of 1-saddle that separate these voids.

This natural association between the various elements of the cosmic web and the elements of the Morse-Smale complex is particularly clear for some morphological elements. To see this, we have to recall that the Morse-Smale complex is focused on the properties of the gradients of the scalar function to determine the partition. With this in mind, we examine this association with respect to voids. A cosmological void is an underdense region, with the center of the void located at a density minimum. Over a period of time, matter streams outward of the void center in all directions. The region of space in which this occurs is precisely the ascending manifold of the minimum. In the case of the cosmic filaments, they are known to be the transport channels of mass to the clusters. Combining this observation with the fact that the destination of the ascending manifold of a 2-saddle are a pair of maxima, and that the gradient of the flow is towards the maxima, justifies the modeling of the cosmic filaments as the ascending manifolds of the 2-saddles. Similarly the clusters in the cosmic web are necessarily located at maxima of the density field. As a result, the nodes of the Morse-Smale complex, which are the positions of maxima are justifiably modeled as the centers of the clusters.

Topology based methods have played a key role in delineating the elements of the cosmic web recently. In the recent past, ideas stemming indirectly (Platen et al. 2007; Aragón-Calvo et al. 2010) or directly (Sousbie 2011) from Morse theory have contributed to the development of structure finders aimed at identifying and delineating the elements of the cosmic web. The watershed transform based void finders

like Watershed Void finder (WVF) (Platen et al. 2007), ZOBOV (Neyrinck 2008), and the Void Identification and Examination toolkit (VIDE) (Sutter et al. 2014a) are based on the segmentation of a terrain using the watershed transform. They identify the voids as the basins of local minima, which is equivalent to the ascending manifolds of the minima in the Morse-Smale complex. The SpineWeb technique (Aragón-Calvo et al. 2010) is an extension of WVF and identifies walls and filaments or the *spine* of the web as the remaining regions which do not belong to the watershed basins. DisPerSE (Sousbie 2011; Sousbie et al. 2011) directly computes the Morse-Smale complex of the density field to identify the clusters, filaments, walls and voids.

Neyrinck (2012) also present a general method of delineating the structures of the web called the ORIGAMI. This is based on the properties of the phase space and and also harnesses Morse theory in its implementation.

It is relevant also to mention here that most of the scalar fields arising in cosmology, like the cosmic density fields, are smooth and continuous, and hence qualify as Morse functions. Throughout this thesis, the analysis of models of cosmic mass distribution assumes that they are Morse functions.

1.7 This thesis

In the previous section, we presented a brief overview of the topological and morphological descriptions involving Euler characteristic and Minkowski functionals. These descriptions, and methods derived from them, have been of key importance in describing and discriminating between the models of the cosmic mass distribution. Subsequently, we introduced the formalisms of homology, Morse theory and persistence and established that they have the potential to describe the topology of the cosmic mass distribution in greater detail. This thesis intends to apply these formalisms, and develop ideas based on them to explore and investigate the models of cosmic mass distribution. The models that we investigate are heuristic, as well as physically more motivated cosmic mass distribution models. The different models mimic different aspects of the cosmic mass distribution.

Random distributions model a variety of phenomena of interest arising in cosmology. The Poisson distribution is used routinely to model shot noise. The deviations of a point distribution from a featureless Poisson distribution indicates the presence of structures in the distribution. Similarly, the Gaussian distribution and Gaussian fields with a flat power spectrum ($n = 0$) are used to model white noise (Bardeen et al. 1986; Gott et al. 2008; van de Weygaert et al. 2011; Park et al. 2013). Another scenario where Gaussian fields are important is in the modeling of the primordial fluctuation field (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986; Park et al. 2013). In view of this, we explore and characterize the topology of the Poisson distribution as well as the Gaussian fields in this thesis. In this context, we note that the topological properties of probability distributions on a manifold, and and the topological properties of random complexes [4] have recently been an active area of research (Adler et al. 2010; Bobrowski & Strom

[4] A *random complex* is a simplicial complex constructed from a given random point distribution The point distribution is sampled from a well behaved probability distribution function.

Borman 2010; Adler et al. 2014; Bobrowski & Kahle 2014; Bobrowski & Mukherjee 2015; Feldbrugge et al. 2015; Wintraecken et al. 2013). For a survey on the topology of random geometric complexes, see Bobrowski & Kahle (2014).

To model the web like patterns of the large scale matter distribution in the Universe, we resort to the Voronoi models (van de Weygaert & Icke 1989; van de Weygaert 1991; Weygaert 2007). The rationale behind the investigation is to identify the topological signatures of the various morphological features of the cosmic web. The Voronoi evolution models are a class of heuristic models for cellular distributions of galaxies that mimics the evolution of the Megaparsec universe towards a web like pattern. They use Voronoi tessellations as a template for distribution of matter and related galaxy population, and its subsequent evolution. The models we use here are considerably sophisticated, and represent a rather realistic depiction of the cosmic web in void-dominated cosmologies (van de Weygaert & van Kampen 1993; Sheth & van de Weygaert 2004; Weygaert 2007; Aragon-Calvo & Szalay 2013). For a more realistic investigation, we also examine the LCDM cosmological simulations. These are fully physical models, based on laws of physics, that trace the formation and evolution of the structures in the cosmic mass distribution.

Another important characteristic of the cosmic mass distribution is its hierarchical nature. To examine this aspect, we use the Soneira-Peebles model (Soneira & Peebles 1978). These are multi-scale fractal models, that are hierarchical in nature. We revisit the hierarchical aspect of the comic mass distribution in the context of a visual exploration and extraction of the filaments in the cosmic web. We define and filter the cosmic filaments by constructing the hierarchical Morse-Smale complex from the LCDM simulations.

1.7.1 Hierarchical characterization of the cosmic mass distribution through persistence

In Section 1.6.3, we briefly describe the formalism of persistence. The most important feature of expressing topology in this fashion in the cosmological context is that it is inherently hierarchical in nature (a full treatment of the topic is presented in Appendix A.5). Persistence topology can naturally account for and identify events that can change the topology of a manifold in a hierarchical build-up of structures, like mergers of isolated objects. As it is established that the structures in the Universe form and evolve hierarchically, a topological description through persistence is naturally tailored towards accounting for the hierarchical nature of the cosmic mass distribution.

There are many scenarios where a hierarchical expression of topology may be relevant. An example is the discrimination between various models of cosmic mass distribution through a topological quantification of the difference in their clustering properties, as recorded through 0-dimensional persistence diagrams. The 0-dimensional diagrams have been noted as indicators of the clustering properties of a manifold (Bobrowski & Mukherjee 2015). Since 0-dimensional diagrams record merger events, any difference arising in them is a direct measure of difference in the clustering properties of the models. In this context, it is important to mention here that the 0-dimensional diagrams should not be confused with the *merger trees* that

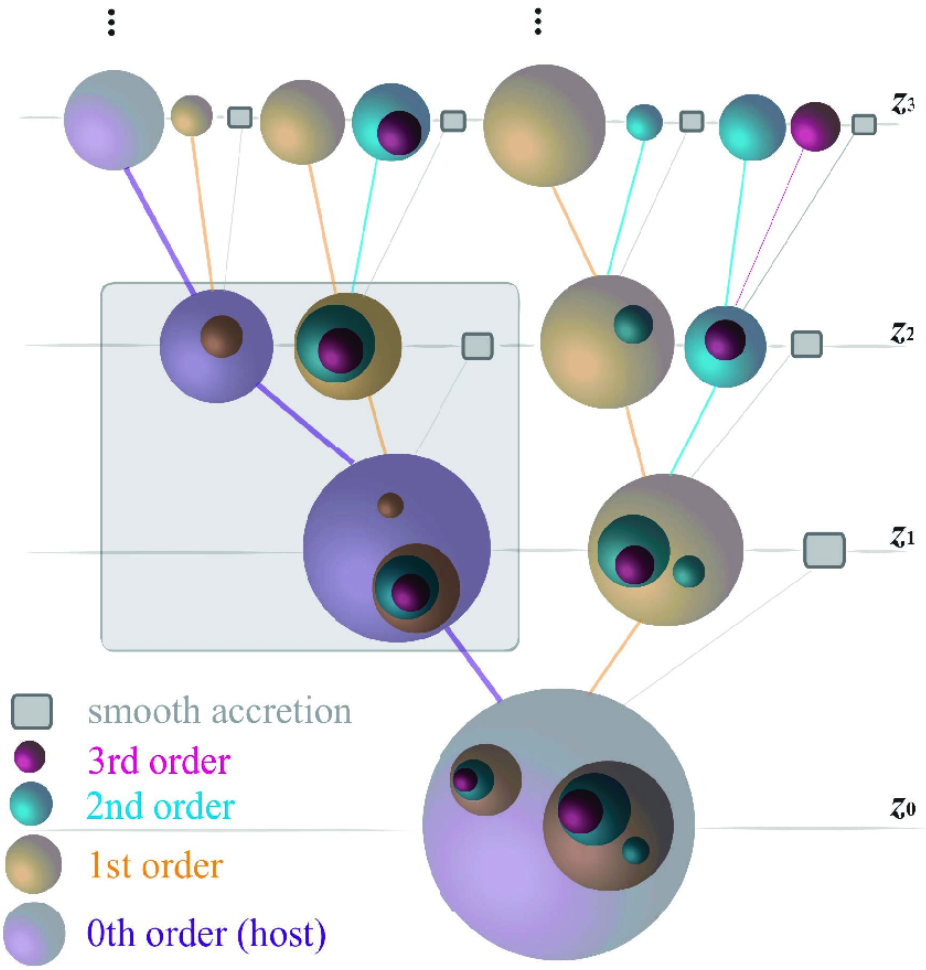


Figure 1.19 Illustration of a typical halo merger tree (Parkinson et al. 2008). From top to bottom along the vertical axis, smaller subhaloes merge with the larger one as time proceeds. This is such that ultimately there is only one large halo with rich substructures left in the end. A 0-dimensional diagram persistence diagram is comparable to the halo merger trees, as the mechanism of merger is intuitively similar. A difference arises from the fact that halo merger trees record mergers as a function of time, and 0-dimensional persistence digram record mergers with changing density threshold. A possible reconciliation may be achieved if a particular parameterization of density with respect to time is assumed.

are traditionally used to represent the time evolution of dark matter haloes through mergers and accretions (Nagashima & Gouda 1997; Shimizu et al. 2002; Parkinson et al. 2008; Behroozi et al. 2013; Jiang & van den Bosch 2014; Lee & Yi 2015). In the cosmological context, the merger trees have been routinely used to depict the time evolution of dark matter haloes through mergers and accretion (see Figure 1.19 for an illustration). The crucial difference between the merger trees and the 0-dimensional

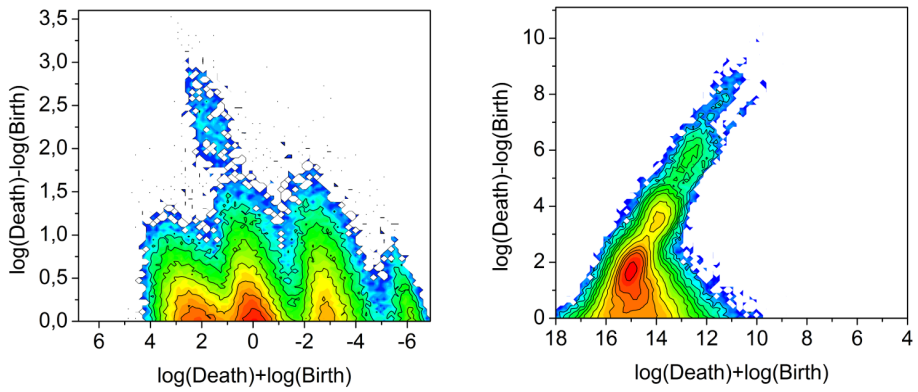


Figure 1.20 Persistence intensity maps for the 0-dimensional holes in highly evolved realizations of the Voronoi kinematic models (left) and the Soneira-Peebles model (right). The map for the Voronoi models, has four distinct features, indicating a presence of four different morphologies, namely clusters, filaments, walls and voids. In the right panel, the maps for the Soneira-Peebles models exhibit a presence of a number of distinct clumps. These clumps are placed at increasingly higher persistence values. These clumps signal the presence of strongly hierarchical distribution.

merger diagrams of persistence is that while merger trees record the merger as a function of time, the persistence merger diagrams record the merger of isolated objects as a function of the density threshold. However, one may attempt to arrive at a relationship between them if some form of dependence is assumed between the evolution of density contrast and time.

Another considerable aspect of persistence is that at no additional cost, we may also extract information about the change in topological behavior of tunnels and voids. Topological behavior of tunnels relates directly to the percolation properties of the model (Barrow et al. 1985; Sahni et al. 1997; Fairall et al. 2005). It will also, be interesting to investigate the topological characteristics of the hierarchical formation of voids for different models of cosmic mass distribution (Sheth & van de Weygaert 2004; Aragon-Calvo et al. 2010; Aragon-Calvo & Szalay 2013; Sutter et al. 2014a). This has the potential to quantify topological signatures arising due to differing prescriptions of dark energy in, for example, the quintessence models.

Motivated by the above observations and discussions, a significant part of this thesis is devoted to a topological characterization of the models of cosmic mass distribution using persistence. Chapter 2 introduces an empirical description of persistence homology for stochastic models of cosmic mass distribution. This is done by defining the *intensity function*, which is the averaged description of persistence over many independent realizations. Chapter 2 also introduces the use of *intensity maps* as pictorial representation of the intensity function.

We analyze the persistence topology of random, featureless distributions, taking up the particular case of Poisson distribution. In the process, we characterize the topology of shot-noise as a benchmark, paving the way for a method to estimate the

contribution of shot-noise in the topological characteristics of a given distribution.

Subsequently, we analyze the persistence topology of the pure Voronoi element models (van de Weygaert & Icke 1989; Icke & van de Weygaert 1991; van de Weygaert 1994). These models are characterized by the presence of predominantly cluster-like, filament-like or wall-like distributions, resembling the different morphologies of the cosmic web (Bond et al. 1996). We show that the dominating presence of different morphologies is reflected in the features of the intensity maps in different dimensions. Specifically, clustered distributions are characterized by high persistent isolated clouds in 0-dimensional diagrams, loopy distributions characterizing highly filamentary patterns are reflected in a similar high persistence cloud in the 1-dimensional intensity maps. Predominantly wall-like distributions, denoting the presence of well formed voids, are characterized by isolated high persistence clouds in the 2-dimensional intensity maps.

We also analyze the persistence topology of the heuristic Voronoi evolution models. The models seek to mimic the evolution of mass distribution in the Universe, gradually progressing from a stage where most of the matter is confined to the field, to a stage where most of the matter is confined to clusters and filaments. We demonstrate that the topological characteristics of the different morphologies present in the distribution – clusters, filaments and walls – are well segregated in the intensity maps.

In the context of hierarchical distributions, we also analyze the persistence topology of the multi-scale fractal Soneira-Peebles model (Soneira & Peebles 1978) in Chapter 2. We demonstrate that the levels of hierarchy are well manifested in the intensity maps.

As an example, Figure 1.20 presents the intensity maps for the 0-dimensional holes in highly evolved realizations of the Voronoi kinematic models (left) and the Soneira-Peebles model (right). The map for the Voronoi models, has four distinct features, indicating a presence of four different morphologies, namely clusters, filaments, walls and voids. In the right panel, the maps for the Soneira-Peebles models exhibit a presence of a number of distinct clumps. These clumps are placed at increasingly higher persistence values. The number of clumps is related to the number of level of hierarchies in the distribution. Therefore, we demonstrate that the intensity maps are able to detect the topological signals of the hierarchical nature of the distribution.

In Chapter 3, we analyze the persistence topology of the Gaussian random fields (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Park et al. 2013). It involves a detailed study of the connectivity characteristics of the density fields modeled by Gaussian random fields. We analyze models specified by power-law power spectra, as well as by the LCDM power spectrum. Keeping model comparison as an important aspect of analysis in mind, we also introduce the difference maps in the chapter. We demonstrate that a combination of intensity, difference and ratio maps express the topological information content in a greater detail compared to the traditional descriptors like the Minkowski functionals.

First, we delve into the persistence topology of 1D Gaussian functions. This is done in order to relate the visual features of the density distribution to the topologi-

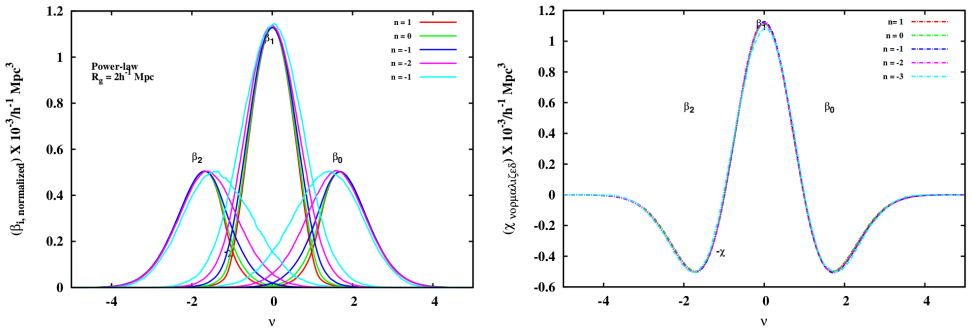


Figure 1.21 The curves of the scaled Betti numbers (left) and the Euler characteristic (right) for Gaussian random fields specified by a power law power spectrum. The curves are drawn for spectral indices $n = 1, 0, -1, -2 \& -3$. The curves are scaled with respect to the amplitude of the curve for the $n = 0$ model. The Betti numbers show a characteristic dependence on the choice of the power-spectrum, unlike the Euler characteristic curve. This means that the Betti numbers are sensitive to the index of the power spectrum, while the Euler characteristic is not.

cal description, and assess how the topological description reflects the features of the density distribution. We establish that the features of the persistence diagram are a direct reflection of the singularity structure of the function.

Having assessed the persistence topology of the 1D functions, we delve into an analysis of the 3D fields. We do this for the power law models and the LCDM model. First, we present the intensity and the difference maps of the power law models and the LCDM model. We demonstrate that the features in the difference maps depend on the choice of the model. The indication is that the the intensity and the difference maps are highly sensitive to the parameters of the model, and may therefore be used to discriminate between various models. This is an important observation in view of model comparison.

We quantify the intensity maps by decomposing them into marginal and cumulative distributions as a function of the mean-density and persistence of the topological holes. In this context, we establish that the distribution functions show a characteristic dependence on the index of the power spectrum. We also show that the marginal distribution of the mean density of the holes follow a near Gaussian distribution. The marginal distribution as a function of the persistence of the holes indicates a Poisson distribution.

1.7.2 Homology and Betti numbers of the models of cosmic mass distribution

Betti numbers have been used to address certain aspects of the cosmic mass distribution concerning the nature of the primordial density field as well as the the topology of large scale structures in the Universe (van de Weygaert et al. 2011; Park et al. 2013). In continuation of the work described in these articles, a part of this thesis seeks to describe in greater detail the topological properties of models of the cosmic mass distribution in terms of the Betti numbers. This involves the Betti characterization of a range of heuristic Voronoi models. These include the single component

Voronoi models, the multi-component Voronoi models and the kinematic Voronoi models. We also present a Betti characterization of the fractal Soneira-Peebles models. Throughout, we demonstrate that the Betti numbers are capture the signals from various morphologies, as well as hierarchies in the relevant scenarios.

Given the importance that Gaussian random fields play in the description of the initial conditions of the cosmic structure formation, we delve into a systematic and detailed analysis of the Betti numbers of Gaussian random fields in Chapter 3. We compare the Betti numbers with the familiar Euler characteristic curves in this context. One of the most crucial observations that we establish is that the Betti numbers show a characteristic dependence on the choice of the power-spectrum, unlike the Euler characteristic curve. This means that the Betti numbers are sensitive to the index of the power spectrum, while the Euler characteristic is not. A visual impression of this may be obtained from inspecting the scaled Betti number and Euler characteristic curves for Gaussian random fields presented in Figure 1.21. The models are characterized by a power law power spectrum. The curves are drawn for spectral indices $n = 1, 0, -1, -2\& -3$. The curves are scaled with respect to the amplitude of the curve for the $n = 0$ model.

A crucial observation is that the topology of the growing superlevel sets of the density field is not strictly either meatball-like, sponge-like or cheese-like. We demonstrate that there are substantial regions of overlap between the various kinds of topologies enumerated above. This is evident from the scaled Betti number curves, which show a substantial overlap across a range of density thresholds. The overlap is the strongest for lower spectral index, and decreases monotonically with increasing spectral index. In contrast, such information may not be available by looking only at the Euler characteristic curves, because their shapes are invariant with respect to the spectral index.

1.7.3 Topology based visualization of Cosmic filaments: Felix

The final part of this thesis introduces and describes a technique for the identification of filaments based on the topological characteristics of the density field. It is based on Morse theory and proceeds by constructing the Morse-Smale complex of the density field. A key aspect of the proposed technique is its interactive nature, and involves a tunable density parameter.

Exploring the filamentary patterns of the cosmic web is challenging because of the large range of the spatial scales and density range it exhibits. A proper characterization should also account for the hierarchical nature of structures, which adds considerable challenge to the task. Though there exist different notions of filaments, the primary evidence relied upon for extraction and analysis is most often visual. It is therefore not surprising that structure finding methods often visually verify results by superimposing the extracted structures upon visualizations of the density field or the particle distribution (Stoica et al. 2005; Aragón-Calvo et al. 2007b; Genovese et al. 2010; Sousbie 2011; Cautun et al. 2013). However the visualization plays a role only after structure extraction process in these methods. We differ in this respect by providing the capability to interact with the structure finding procedure and extract structures that are visually relevant. The work emanating from this thesis in-

creases, e.g. the scope of investigating, in a more systematic fashion, the dependence of formation, evolution and properties of galaxies on the surrounding large scale environment.

Density range based modeling of filaments

The distinction between noise and significant structures is often ill-defined, and at occasions noise may be confused with genuine structures in the hierarchically evolved mass distribution. This problem is more pronounced when one studies the properties of tenuous filaments and walls in low density void-like regions. For the understanding of the formation and evolution of galaxies in such regions, we need to assess the possible dependence of galaxy and halo properties on the morphology and density of the local environment (Jones et al. 2010; Tempel et al. 2013; Cautun et al. 2014). This must be based on the successful extraction of filaments in low density regions and the correct identification of galaxies associated with them. In view of this, we describe Felix[5], a topology based visual query framework to extract filamentary structures from a hierarchy of Morse-Smale complexes of the density field. The filaments in Felix are parameterized by the density values of the maxima and the 2-saddle that define them. In the software, we incorporate an interactive handle to specify the density range of the maxima and the saddles. This feature allows the user to concentrate on and probe structures in specific density regimes. It also offers the user the possibility to switch between the various regimes of possible interest in real time.

As an example, we investigate the nature of filaments in three different density regimes from the ΛCDM simulations. The first concerns filaments in the high density regions around compact dense clusters, which are known to function as the transport channels along which matter moves into the clusters. The second regime concerns the tenuous low-density filaments found in low-density void regions. The third regime concerns the filaments that stretch from cluster like regions all the way down to void like regions. Figure 1.22 presents three classes of filaments extracted from a cosmological simulation. Panel (a) shows the particle distribution, along with a volume rendering of the density field. Panel (b) presents the Filaments within cluster like regions. The highlighted region shows the retention of intricate topological structures within a large cluster region. Panel (c) presents the filaments within void like regions. Shown in the inset is a cluster like region within which filamentary structures are filtered away as desired by the query to the framework. Panel (d) presents the filaments that stretch from cluster like regions all the way down to void like regions. The highlighted region shows a single filament passing through the large cluster like region with intricate topological details filtered out by the framework.

Testing and comparing Felix

In order to test the efficacy of Felix, we compare it with a few structure finders. We do this by developing a semi automatic structure finder, based on Felix, that classifies galaxies as cluster/filamentary or not. Using the Voronoi Kinematic model

[5]The name Felix is formed from an abbreviation of *Fil*ament *ex*plorer.

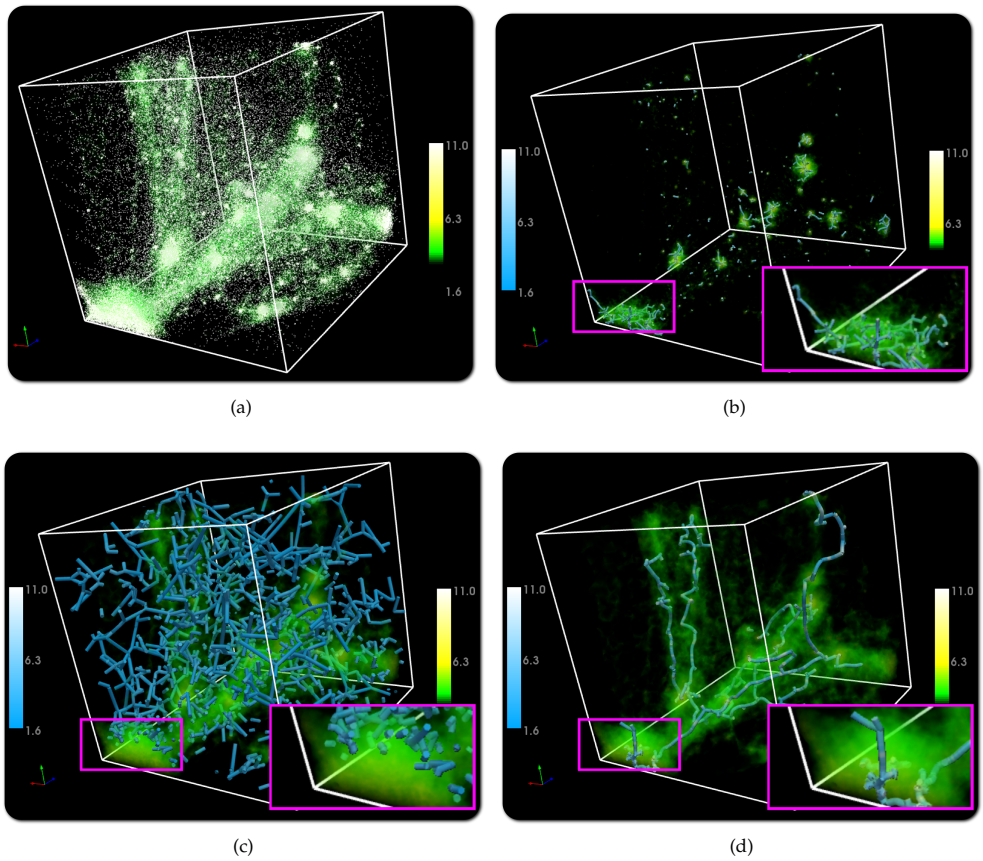


Figure 1.22 Three classes of filaments extracted from the Cosmogrid dataset. (a) Particle distribution shown along with a volume rendering of the DTFE density. (b) Filaments within cluster like regions extracted with parameters $[S_b, S_e] = [10^5, 10^8]$ and $[M_b, M_e] = [10^8, 10^{12}]$. The highlighted region shows the retention of intricate topological structures within a large cluster region. (c) Filaments within void like regions extracted with parameters $[S_b, S_e] = [10^0, 10^5]$ and $[M_b, M_e] = [10^0, 10^5]$. Shown in the inset is a cluster like region within which filamentary structures are filtered away as desired by the query to the framework. (d) Filaments that stretch from cluster like regions all the way down to void like regions using parameters $[S_b, S_e] = [10^3, 10^{11}]$ and $[M_b, M_e] = [10^9, 10^{11}]$. The highlighted region shows a single filament passing through the large cluster like region with intricate topological details filtered out by the framework.

as a benchmark, we demonstrate that we are able to recover the classification with high efficiency. We compare Felix with a few other popular methods, namely DisPerSE (Sousbie 2011), MMF/Nexus (Aragón-Calvo et al. 2007b; Cautun et al. 2013) and SpineWeb (Aragón-Calvo et al. 2010). We demonstrate in the process that we perform equally good, if not better, in most conditions. This includes models which are not very highly evolved and hence lack a strong morphological segregation.

We pay a particular attention to the comparison between DisPerSE (Sousbie 2011; Sousbie et al. 2011) and Felix. These are closely related structure finders, as both use the Morse-Smale complex of the log-density field and involve feature extraction

from it. DisPerSE simplifies the Morse-Smale complex using Topological Persistence. It defines significant features as only those that remain unsimplified using the user defined significance threshold. It ignores the density range characteristics of the extracted features. A significant consequence is that filaments within void-like regions and cluster like regions are ignored/simplified away. If they are retained, then the mixing of features causes visual clutter. Furthermore, the significance parameter selection is a fixed constant and visual interaction plays no role in its selection. In contrast, given the ubiquity of filaments in various density regimes, Felix allows for density ranged based probes into filaments, within clusters and voids. Another difference is that Felix uses simplification only for noise removal and not feature identification.

We construct an experiment to demonstrate the consequences of not correlating the density characteristics for filament extraction. Using the Voronoi models as a test, we show that tuning the significance parameter, as implemented in DisPerSE, is not a sufficient mechanism to extract the desired filaments. Our findings suggest that a structure identification strategy based on a direct simplification procedure of the Morse-Smale complex should be applied with care to the density regimes being studied. By comparing the true and false detection rate profiles we confirm that, using Felix, we extract filaments that are spatially more proximal to the cluster and filament particles in the Voronoi Kinematic datasets, as compared to DisPerSE. We also show that in extreme situations the false detections from DisPerSE have value larger than 1. This is indicative of an over-detection of filaments. This is potentially cumbersome for the analysis of genuine cosmological simulations and observational surveys. In more complex realistic circumstances, cosmic structure involves features over a wide range of densities and scales and structural morphologies that are not as well separated as in the simpler Voronoi models.

In a separate experiment, we demonstrate the exploration of filaments within high-density cluster like regions and low-density void like regions. Such a delineation, coupled with the visual exploration process, is not possible using DisPerSE.

1.7.4 Outline of this thesis

In Chapter 2, we present a formal discourse on the concepts of Morse theory, homology and persistence homology. Based on this we introduce the *intensity function* and *intensity maps* as empirical probabilistic description of persistence for stochastic processes. We then analyze a few heuristic models of cosmic mass distribution in terms of Betti numbers and persistence.

We proceed to analyze the (persistence) homology of Gaussian random fields in Chapter 3. In the context of model discrimination, we introduce the use of the difference maps. We also analyze the Betti numbers of Gaussian random fields and show that they contain strictly extra information compared to Euler characteristic and genus. In this context, we arrive at an expression of the scaling properties of Betti numbers.

In Chapter 4, we exploit the geometry of the manifolds associated with the critical points, in combination with persistence, to develop an interactive software to identify the cosmic filaments. This harnesses the hierarchical nature of persistence, and

leverages this to identify structures in the web in a naturally hierarchical fashion. We compare the method to some of the existing procedures through benchmark models, and demonstrate that we fare equally good, if not better, in almost all the cases. The procedure is also naturally suited for additional features like noise removal, based on topological simplification of the manifold. We propose the software as a rich tool towards investigating the filamentary structures of the cosmic web.

Chapter 5 concludes this thesis with a summary of the main observations and conclusions of this thesis, and an outline of some potential problems of relevance that can be studied through techniques described here.

2

On the Betti of the Universe, and her persistence

— Pratyush Pranav—

Every passing hour brings the Solar System forty three thousand miles closer to Globular Cluster M13 in Hercules and still there are some misfits who insist that there is no such thing as progress.

– *Kurt Vonnegut, The Sirens of Titan*

2.1 Introduction

The Megaparsec scale distribution of matter revealed by galaxy surveys features a complex network of interconnected filamentary galaxy associations. This network, which has become known as the *Cosmic Web* (Bond et al. 1996), contains structures from a few megaparsecs up to tens and even hundreds of megaparsecs of size. The cosmic mass distribution displays a wispy web-like spatial arrangement consisting of dense compact clusters, elongated filaments, and sheet-like walls, amidst large near-empty voids, with similar patterns existing at earlier epochs, albeit over smaller scales. The multi-scale nature of this mass distribution, marked by substructure over a wide range of scales and densities, has been clearly demonstrated. Its appearance has been most dramatically illustrated by the recently produced maps of the nearby cosmos, the 2dFGRS, the SDSS, and the 2MASS redshift surveys (Colless et al. 2003; Tegmark et al. 2004; Huchra et al. 2005).

The vast Megaparsec Cosmic Web is one of the most striking examples of complex geometric patterns found in nature, and certainly the largest in terms of sheer size. Computer simulations suggest that the observed cellular patterns are a prominent and natural aspect of cosmic structure formation. These structures form via the mechanism of gravitational instability (Peebles 1980), which is the standard paradigm for the emergence of structure in our Universe (Springel et al. 2005). According to the *gravitational instability scenario*, cosmic structure grows from tiny primordial density and velocity perturbations. The evidence provided by the temperature fluctuations in the cosmic microwave background (Smoot et al. 1992; Bennett et al. 2003; Spergel et al. 2007; Komatsu et al. 2010; Planck Collaboration et al. 2013b) suggests that the character of the perturbation field is that of a homogeneous and isotropic spatial Gaussian process. Theories of the early universe predict such primordial Gaussian perturbations in the gravitational potential, as they are a natural product of an early inflationary phase of our Universe.

Since dark matter is the dominant component of the cosmic matter distribution, its nature is of utmost importance in determining the direction that structure formation in the Universe takes. Depending on whether the dark matter is cold or hot, structure formation proceeds via either the bottom-up (Peebles 1980) or the top-down mechanism(Zel'dovich 1970). In the bottom up scenario, the smallerer objects like galaxies and groups of galaxies are the the first to form. Subsequently, they merge to form larger and larger structures. In the top-down scenario, the large scale structures are the first ones to form. Subsequently, the matter inside them fragments or *coagulates* to form smaller structures. The primordial Gaussian density field is probabilistically characterized by its power spectrum. The presence of dark matter modifies the form of the primordial spectrum at a later epoch through a transfer function which depends upon the nature of dark matter. This happens because of coupling of fourier modes in the non-linear regime of gravitational amplification. Observational measurements of the power spectrum of matter distribution in the Universe indicate that the bottom-up scenario is the favored one (Planck Collaboration et al. 2013a; Tegmark et al. 2004; Komatsu et al. 2009).

It has remained a major challenge to characterize the structure, geometry, and topology of the Cosmic Web. The overwhelming complexity of both the individual

structures as well as their connectivity, the lack of structural symmetries, the intrinsic multi-scale nature of the web, and the wide range of densities in the cosmic matter distribution has prevented the use of simple and straightforward instruments. Many attempts to describe and quantify, let alone identify, the features and components of the Cosmic Web have therefore been of a rather heuristic nature. Measures like the two-point correlation function, which has been the mainstay of many cosmological studies over the past forty years (Peebles 1980), are not sensitive to the spatial complexity of patterns in the mass and galaxy distribution.

A key aspect of the complexity of the Cosmic Web is the connectivity of the various structural components. Topology is the branch of mathematics that addresses issues of shape and connectivity. The first cosmological studies that focused on topological characteristics of the cosmic mass distribution evaluated and analyzed the genus and the Euler characteristic of the cosmic mass distribution. Gott and collaborators (Gott et al. 1986; Hamilton et al. 1986) studied the genus as a function of density threshold. Later, more discriminative topological information became available with the introduction of Minkowski functionals (Mecke et al. 1994; Schmalzing & Buchert 1997). However, nearly without exception these studies had a largely heuristic character, driven mainly by the urge to find additional useful measures for discriminating between different cosmologies.

There is ample motivation to extend topological studies beyond just discriminating between various cosmologies. For example, an interesting aspect that could have an answer through topological studies is the delineation of signals of primordial non-Gaussianities from the non-Gaussianities developing due to gravitational amplification in the later non-linear epochs (Dominik & Shandarin 1992). Similarly, questions on if one can differentiate between signals coming from various elements of the cosmic web based on their topological characteristics is an important one. Perhaps one of the most motivating facts for developing the topological characterization of the cosmic matter distribution beyond the existing techniques is the recognition that the matter distribution is distinctly hierarchical in nature. In view of this, a formalism capable of expressing topology in a hierarchical fashion is an interesting and powerful alternative to the existing methods.

For a more profound understanding of the topological structure of the cosmic mass distribution, we have to relate the topological measures to the underlying singularity and connectivity characteristics. There is a wealth of information to be gained from a systematic analysis of the singularity structure of a field. This is because the singularities, or the critical points, determine the overall topological structure of the field. The basis of this is rooted in Morse theory (Milnor 1963), which seeks to study the change in topological properties of a manifold through a scalar function defined on it. With a slight twist, this change in topology can be expressed in a hierarchical nested fashion, as is achieved in the formalism of *persistence* (Edelsbrunner et al. 2002; Edelsbrunner & Harer 2010). Informally, this relates to the *birth* and *death* of a topological hole, the dimensions of which are determined by the index of the critical point associated with its birth and death. See Chapter 1, Section 1.6.3 and Appendix A for a detailed exposition on the procedure.

A more formal definition of topological holes is through algebraic topology, specif-

ically called *Homology*. The rank of the p-dimensional *homology group* $\mathbb{H}_p$, denoted by the *Betti number* β_p (Betti 1871), counts the number of independent p-dimensional cycles in a manifold. These *cycles* bound the holes of the manifold. For a detailed exposition on holes and homology, refer to Appendix A.

However, Betti numbers attach the same weight to all topological holes. This is not desirable : in the context of cosmological data sets, almost all data is discretely sampled – particles in N-body simulations and discrete galaxies in observations. This introduces erroneous topology at the smallest scales, arising from discrete sampling. Persistence deals with this by formalizing topology as a hierarchical concept. In this formalism, one records the creation (birth) and destruction(death) of topological holes, as one sweeps from the highest to the lowest function values. The topological holes form and destroy only when crossing the critical points of the function. The absolute difference in the function value between the death and birth of topological holes quantifies its life-span or *persistence*. While the definition and the meaning of "noise" is case specific, in general, low-persistence holes are more likely to be topological noise, while high-persistence holes quantify real signals.

The goal of this chapter is to introduce concepts derived from homology, Morse theory and persistence and analyze heuristic models that mimic certain aspects of the mass distribution in the Universe. To this end we introduce the concept of intensity maps as an empirical description of persistence diagrams. The ensemble average of persistence diagrams emanating from realizations of stochastic processes converge over many realizations to a stable function. The intensity maps also reveal the hidden structures in the the persistence diagrams, which is not apparent in the scatter plot representation. Specifically, they are able to highlight specific features of the mass distributions in the models by virtue of being able to differentiate between topological signals coming from different structural components, as for example in the Voronoi evolution models (van de Weygaert 1991). They are also able to resolve the levels of hierarchies in fractal distributions remarkably well, as for example in the Soneira-Peebles model (Soneira & Peebles 1978).

In order to explore and understand what a topological description through persistence and homology has to offer, we analyze a few different heuristic models of mass distribution in the Universe in this chapter. We briefly discuss the topological concepts essential to the analysis, relegating a detailed description to Appendix A. In Section 2.4, we seek to analyze the topology of random, featureless distributions, taking up the particular case of Poisson distributed particles through persistence intensity maps and the Betti numbers. In Section 2.5, we analyze the topology of the pure Voronoi element models. These models are characterized by the presence of either a cluster-like, filament-like or wall-like distributions, resembling the different morphological elements of the cosmic web (Bond et al. 1996). Section 2.6 analyzes the topology of the multi-scale fractal Soneira-Peebles model (Soneira & Peebles 1978). The analysis of this model is motivated by the fact that it can be tuned to match the observed angular distribution of galaxies in the sky. In Section 2.7, we analyze the topology of the heuristic Voronoi evolution models. The models seek to mimic the evolution of mass distribution in the Universe, gradually progressing from a stage where most of the matter is confined to the field, to a stage where most of the matter

is confined to clusters and filaments. Section 2.8 presents an exposition on a potential method to separate the contribution of shot noise, arising from discrete sampling, from the real topological signals. Finally, in Section 2.9, we analyze the topological characteristic of a single Voronoi evolution model, where the density is computed through different prescriptions. To this end, we construct the density fields from the underlying discrete particle distribution using the Delaunay Tessellation Field Estimator (DTFE) (van de Weygaert & Schaap 2009a), and the density computed on a regular grid using Smoothed Particle Hydrodynamics (SPH) (Hernquist & Katz 1989; Monaghan 2005). This is motivated by the urge to understand how the different density estimators affect the topological characteristics of the model. We conclude the chapter in Section 2.10.

2.2 Models

In this study, we seek to analyze the homology of cosmological density fields. The mass distribution in the Universe is described by the density perturbation field,

$$\delta(\vec{x},t) \;=\; \frac{\rho(\vec{x},t)-\rho_u(t)}{\rho_u(t)}\,, \tag{2.1}$$

which describes the fractional over- or underdensity at position $\vec{x}$ with respect to the universal mean cosmological density $\rho_u(t)$. By the definition above, $\delta \geq 1$.

The models used in the topological analysis are heuristic, but representative of the certain aspects of the cosmic mass distribution. We begin by analyzing the topology of the featureless Poisson distribution. This analysis is motivated by an urge to understand the topological characteristics of the imprints that the ubiquitous shot-noise leaves on the cosmological data sets. Thereafter, we move on to the analysis of Single element Voronoi models, to enhance our understanding on the topological characteristics of the various morphological elements of the cosmic web (refer to Appendix D for the details of the model). The characteristics of hierarchical distributions are analyzed through the multi-scale Soneira-Peebles model (refer to Appendix E for the details of the model). Finally, we analyze the Voronoi evolution models (described in Appendix D) in a bid to understand the topological properties of dynamically evolving mass distributions, in keeping with the fact that that mass distribution in the Universe evolves from a rather featureless distribution at the earlier epochs to a distribution with distinct morphological features at the present epoch in the cosmic timeline.

2.2.1 Level sets, superlevel sets and filtration

When assessing the mass distribution by a continuous density field, $f(\mathbf{x})$, a common practice is to study the sublevel or superlevel sets of the mean square fluctuation smoothed on a given scale R_s, which is given by

$$\sigma^2(R_s) = \int d^3\,kW^2(kR)P(k), \tag{2.2}$$

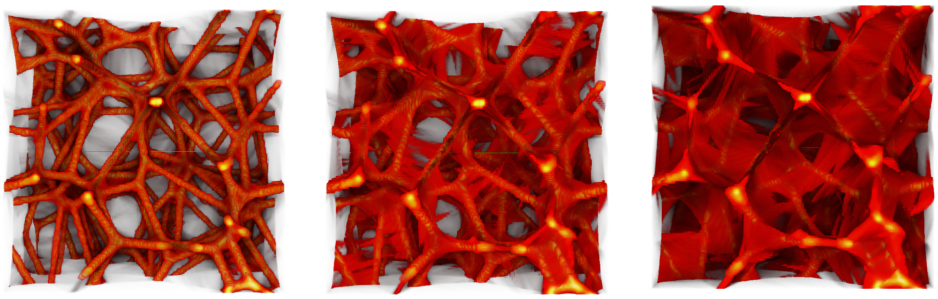


Figure 2.1 Density rendering of the superlevel set of the pure filamentary models. From top to bottom: three snapshots for growing superlevel sets.

where $W(x)$ is the smoothing kernel. The *superlevel sets* of this field are defined as the regions

$$\mathbb{M}_\nu = \{\mathbf{x} \in \mathbb{M} \mid f_s(\mathbf{x}) \in (f_\nu, \infty]\} \tag{2.3}$$
$$= f_s^{-1}(-\infty, f_\nu]. \tag{2.4}$$

In other words, they are the regions where the smoothed density is greater than or equal to the threshold value $f_\nu = \nu\sigma_0$, with σ_0 the dispersion of the density field. A typical example of superlevel sets of a density field is that shown in figure 2.1.

When addressing the topology of a mass or point distribution, a rich source of information is the topological structure of a filtration. Given a space $\mathbb{M}$, a *filtration* is a nested sequence of subspaces:

$$\emptyset = \mathbb{M}_0 \subseteq \mathbb{M}_1 \subseteq \ldots \subseteq \mathbb{M}_m = \mathbb{M}. \tag{2.5}$$

The nature of the filtration depends, amongst others, on the representation of the mass distribution. It is the evolving topology as we pass through the filtration sequence which represents a rich source of information on the topological complexity of the field.

Figure 2.1 provides a telling illustration of a density-defined filtration of a weblike spatial pattern. It concerns a model of the cosmic web consisting exclusively of filaments (see appendix D for a detailed description). It shows a sequence of three growing superlevel sets of the weblike density field, along a sequence of decreasing density threshold. The top panel corresponds to the highest density threshold. It reveals the high density regions that outline the underlying skeleton. The additional panels reveal complementary information on the manner in which matter has distributed itself over the various structural components, revealing how the lower density mass elements connect up and fill in the interstitial regions of the network. The illustration shows how the sequence of filtration steps establish the connectivity of the cosmic mass distribution, and entails its topological structure.

2.2.2 Piecewise Linear scalar fields.

In many practical circumstances, whether it concerns the spatial distribution of galaxies in redshift surveys or particles in cosmological N-body simulations, we are dealing with data sets consisting of discrete particle positions. There are various ways in which the topology of such a discrete particle data set can be analyzed. One option is to define a filtration on the point distribution itself. The most direct way to achieve this is that via a simplicial complex generated by the point distribution. Well-known examples are that of the alpha-complex (Edelsbrunner & Mücke 1994; Edelsbrunner & Harer 2010) and the Cech complex (see Edelsbrunner & Harer 2010), invoking the distance function and a corresponding distance parameter to define the filtration.

In our study we follow a different approach. The topological analysis in our study is based on a density value-based filtration of a piecewise linear density field. The latter is computed from the discrete particle distribution itself. The filtration consists of density value superlevel or sublevel sets.

The determination of a piecewise linear density field from a discrete particle distribution involves a few key steps. The first step involves an estimate of the density at each of the sample points. Usually, the particles define the point sample, but in principle one may define alternatives. The second step involves the determination of a triangulation on the basis of the point sample. In each triangle m (2D) or tetrahedron m (3D) of the triangulation, the gradient $\widehat{\nabla f}\big|_m$ can be uniquely determined from the $D+1$ vertices.

For a sample of N points, with density value estimates $f(\vec{x}_j)$ ($j = 1, \ldots, N$), the density value $\widehat{f}(\vec{x})$ at location $\vec{x}$ is uniquely determined from the density gradient $\widehat{\nabla f}\big|_m$ of the triangle/tetrahedron m in which it is located, and the density value at one of its vertices, $\vec{x}_i$),

$$\widehat{f}(\vec{x}) \,=\, \widehat{f}(\vec{x}_i) \,+\, \widehat{\nabla f}\big|_m \cdot (\vec{x} - \vec{x}_i)\,. \tag{2.6}$$

One key element of a procedure to construct a linear piecewise density field is the nature of the estimate of the density at each sample point. A second key element is the nature of the triangulation. In most of our results, we use the Delaunay Tessellation Field Estimator, DTFE (Schaap & van de Weygaert 2001; van de Weygaert & Schaap 2009b; Cautun & van de Weygaert 2011). It is based on local density estimates determined from the volume of the corresponding contiguous Voronoi cell or *star*, and the use of the Delaunay tessellation as triangular interpolation mesh. A summary of the technical details of DTFE can be found in appendix F.0.6.

2.3 Topology

In this section, we introduce the topological concepts we use to analyze particle distributions, through a running example. We relegate the formal definition of these concepts to the Appendix A. The main new methods for cosmological applications are Betti numbers and persistence, which we will relate to the more traditional notions of Minkowski functional, Euler characteristic, and genus.

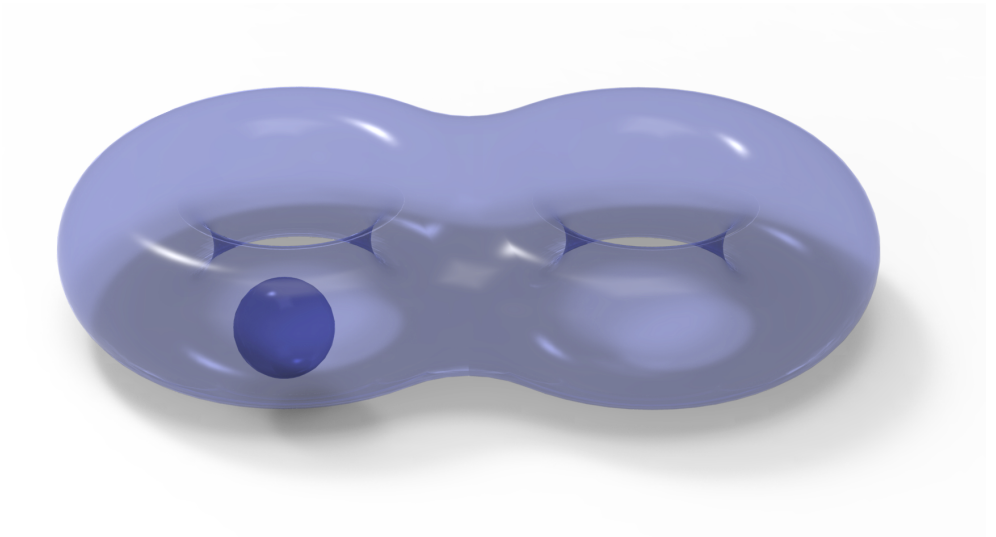


Figure 2.2 A solid double-torus with a bubble.

2.3.1 Running example

Let $\mathbb{M}$ be a solid double-torus with an empty bubble, that is: a double-donut with a small void inside; see Figure 2.2. Its boundary, denoted as $\partial\mathbb{M}$ consists of two surfaces: a double-torus on the outside and a sphere bounding the bubble.

The *Minkowski functionals* are the volume of $\mathbb{M}$, the area, the total mean curvature, and the total Gaussian curvature of $\partial\mathbb{M}$. These are geometric properties, but they are not independent of the purely topological concepts we will introduce next.

The *Euler characteristic* is the alternating sum of simplex numbers needed to triangulate a manifold. Applied to $\partial\mathbb{M}$, the number of vertices minus the number of edges plus the number of triangles needed to triangulate the double-torus gives -2, and for the sphere we get $+2$. It follows that the Euler characteristic of $\partial\mathbb{M}$ is $\chi = 0$. There are many other 2-dimensional manifolds that have the same Euler characteristic, the torus being one, the union of two tori being another.

The Gauss-Bonnet Theorem relates the Euler characteristic with the total Gaussian curvature, stating that for a connected surface, the latter is equal to $2\pi\chi$. Indeed, the total Gaussian curvature of the sphere is 4π, no matter how large it is, and the Euler characteristic of the same is 2. The *genus* of $\partial\mathbb{M}$ is 2, namely 2 for the double-torus plus 0 for the sphere. For a connected surface, the genus equals 1 minus half the Euler characteristic. More generally, the genus of a 2-manifold that is the union of disjoint surfaces is therefore

$$g = \sum_i g_i = \sum_i \left(1 - \frac{\chi_i}{2}\right) = \#\text{components} - \left(\frac{\chi}{2}\right), \tag{2.7}$$

where we write χ_i and g_i for the Euler characteristic and the genus of the i-th component. The reader may check that this relation holds for $\partial\mathbb{M}$. We get a refinement of the concepts by introducing *Betti numbers*. Formally, they are ranks of homology groups, one for each dimension (more on homology and homology groups later). We have

$$\begin{aligned} \beta_0 &= \text{\#components,} \\ \beta_1 &= \text{\#independent loops,} \\ \beta_2 &= \text{\#independent closed surfaces.} \end{aligned} \tag{2.8}$$

For $\partial\mathbb{M}$, we have $\beta_0 = 2$, $\beta_1 = 4$, $\beta_2 = 2$. Indeed, we have two components and two closed surfaces: the double-torus and the sphere. To see the four loops, draw one around each hole of the double-torus and another one around each handle. We get the Euler characteristic by taking the alternating sum: $\chi = \beta_0 - \beta_1 + \beta_2$, which for $\partial\mathbb{M}$ gives 0, as required.

Suppose now that $\mathbb{M}$ is the portion of the Universe at which the local density exceeds some threshold, ν. What if we decrease ν by some small but positive amount? Decreasing the threshold enlarges the portion at which the density threshold is exceeded. It may be that the bubble fills up. Assuming that nothing else changes, $\partial\mathbb{M}$ is now a double-torus, with $\beta_0 = 1$, $\beta_1 = 4$, $\beta = 1$. The sphere and the bubble have gone.

There is a wealth of information in this game of birth and death that unfolds when we continuously decrease the density threshold. Harvesting this information is perhaps the most important new idea this paper brings to the study of the cosmic web. A formal description of this process requires the homology groups, of which the Betti numbers are the ranks, and the maps between these groups that are induced by the inclusions of one body in the next. The reader who is not inclined to learn about these algebraic concepts can still follow the text and the results in this paper, basing her intuition on the closely related but more intuitive birth and death game we see in Morse theory (refer to Appendix A.4, for an exposition on birth and death of topological holes and the connection to Morse theory). It suffices to mention that the a compact representation of the topological changes across the whole range of function values on a manifold is captured by the *persistence diagrams* (Edelsbrunner & Harer 2010). It is a collection of dots in $\mathbb{R}^2$, where the dots have the same units as the scalar function defined on the manifold. The usual practice is to represent the birth value of a feature on the horizontal axis, and the corresponding value of density threshold for death on the vertical axis. There are $0-, \ldots, (d-1)-$ dimensional diagrams for a d-dimensional manifold. We use the persistence diagrams to arrive at the concept of *intensity function* and *intensity maps* as empirical probabilistic description of persistence homology. This is a novel contribution of this thesis, and we describe it next.

2.3.2 Persistence intensity function and intensity maps

This chapter concerns itself with the topology of stochastic point processes, and density field computed on them. In the context of the Universe, both the cosmic microwave background and the density distribution in the Universe are examples of spatial stochastic processes. It is a universal property of stochastic processes that the expectation value of the quantities defined on them converge over many realizations. Our conjecture is that this must also be true for the birth-death events, as reflected in the persistence diagrams, if averaged over many realizations. While a rigorous attempt at deriving a probabilistic and statistical description of persistence topology is beyond the scope of this chapter, we provide an empirical description and test, as proof of the hypothesis, by introducing the intensity maps.

We are interested in the statistical description of persistence diagrams, as an average over many realization, of the stochastic process f. To this end, we construct the *intensity maps*, which is the function $p : \mathbb{R}^2 \to \mathbb{R}$ in the *mean density-persistence* plane, whose integral over every region $R \subset \mathbb{R}^2$ is the expected number of points in R, $\langle N_{tot} \rangle$. Note that $\langle N_{tot} \rangle$ is a representative of the *total intensity* of the map. We discretize the intensity map into a number of regular grid-cells in the plane, and define the intensity function, for the grid-cell (i,j) as

$$I_{ij} = \frac{\langle N_{ij} \rangle}{\langle N_{tot} \rangle}. \tag{2.9}$$

The total *intensity* of the maps is proportional to the average number of total dots in the persistence diagrams. For each grid cell, the intensity function represents the fraction of the total intensity of the map. Since the intensity in each bin is normalized by the total intensity of the map, the integral of the intensity function over $\mathbb{R}^2$ always evaluates to 1, irrespective of the model in question. In the limit of the size of the grid-cells going to zero, the discretized intensity function approximates the probability density function. At this point, we only have empirical evidence that if f arises from a stochastic process and is is tame (all the derivatives well defined), the intensity maps are well defined . As we will show shortly, the intensity function and intensity maps are highly sensitive to the parameters of the model, and capture local variations in topology across the whole range of function value. As such, we propose their use to characterize and discriminate between various models.

Our preferred visual presentation of a diagram is averaged over a number of realizations of the same random experiment; see Figure 2.7, which shows the plots for the data generated as described in Section 2.4. To construct it, we superimpose the diagrams of the different realizations, we discretize $\mathbb{R}^2$ using a grid of 100-by-100 squares, and we form the histogram by counting the points in each square. The result is a real-valued function on the plane, which we denote as the *averaged persistence diagram* or the *intensity map* of the diagram.

2.3.3 Running example : persistence homology of a triangle

In this section, we illustrate the construction of filtration and the concept of boundary matrix. The reduced boundary matrix encodes the persistence diagram. We take a

triangle as our input simplicial complex.

Filtration

We assume there is a function defined on the simplices that constitute the triangle. The function is such that it induces an ordering of the simplices, from the lowest to the highest dimension. Figure 2.3 depicts such an ordering and the order in which the simplices appear in the filtration. We examine the filtration now, while simultaneously keeping track of the birth and death events.

First the vertex $\langle 1 \rangle$ appears in the filtration. This corresponds to the birth of a 0-dimensional hole, or an isolated object. Subsequently, vertices $\langle 2 \rangle$ and $\langle 3 \rangle$ appear, in that order, taking the number of isolated objects to 3. order. Thereafter, the edge $\langle 4 \rangle$ appears, merging the vertices $\langle 1 \rangle$ and $\langle 2 \rangle$ into a single component. We have a death of a 0-dimensional hole here. According to elder rule, the component that forms early lives, and the younger component dies. In other words: the edge $\langle 4 \rangle$ kills the vertex $\langle 2 \rangle$, and $\{\langle 2 \rangle, \langle 4 \rangle\}$ form a birth-death persistence pair in the corresponding to a 0-dimensional hole. Thereafter comes edge $\langle 5 \rangle$, merging the vertex $\langle 3 \rangle$ with the connected component $\langle 1 \rangle$ (note that, since $\langle 2 \rangle$ is dead, the connected component resulting from the merger of $\langle 1 \rangle$ and $\langle 2 \rangle$ has the same index as $\langle 1 \rangle$).

The first topological hole in 1-dimension is born when the edge $\langle 6 \rangle$ appears in the filtration. this completes the triangle, forming a loop. This 1-dimensional hole dies when the face of the triangle appears in the final phase of the filtration, patching up the loop that had formed due to the introduction of the edge $\langle 6 \rangle$. In other words, $\{\langle 6 \rangle, \langle 7 \rangle\}$ form a birth-death persistence pair in 1-dimension.

In summary, there are three birth-death pairs in the filtration of the triangle : two corresponding to isolated components – $\{\langle 2 \rangle, \langle 4 \rangle\}$ & $\{\langle 3 \rangle, \langle 5 \rangle\}$, and one corresponding to the loop – $\{\langle 6 \rangle, \langle 7 \rangle\}$.

From the point of view of need to construct the boundary matrix, we also enumerate the simplices and their boundaries here. The boundary of the edges constitutes of the vertices – for example, the boundary of the edge $\langle 4 \rangle$ consists of the vertices $\langle 1 \rangle$ and $\langle 2 \rangle$. The boundary of the triangular face $\langle 7 \rangle$ consists of the edges $\langle 4 \rangle$, $\langle 5 \rangle$ and $\langle 6 \rangle$.

Boundary matrix and its reduction

We construct the boundary matrix, ∂ of the filtration of the triangle (refer to Appendix B.5 for a detailed account on boundary matrices and the algorithm for computing persistence). Since the number of simplices in the filtration is 7 (3 vertices, 3 edges, and 1 face), the size of the boundary matrix is 7×7. If the simplex i is in the boundary of the simplex j, the (i, j)-th element of the matrix is 1. All other elements are 0. We reduce the boundary matrix to R, using Algorithm 2, to the form detailed in Appendix B.5. Figure 2.4 illustrates this operation in the form of the matrix multiplication notation $R = \partial \cdot V$, where R and ∂ are the reduced matrix and the original boundary matrix respectively. One may verify that the shaded entries in the ∂ matrix of Figure 2.4 indeed correspond to the simplices of the triangle, and its boundary (Figure 2.3 and Section 2.3.3).

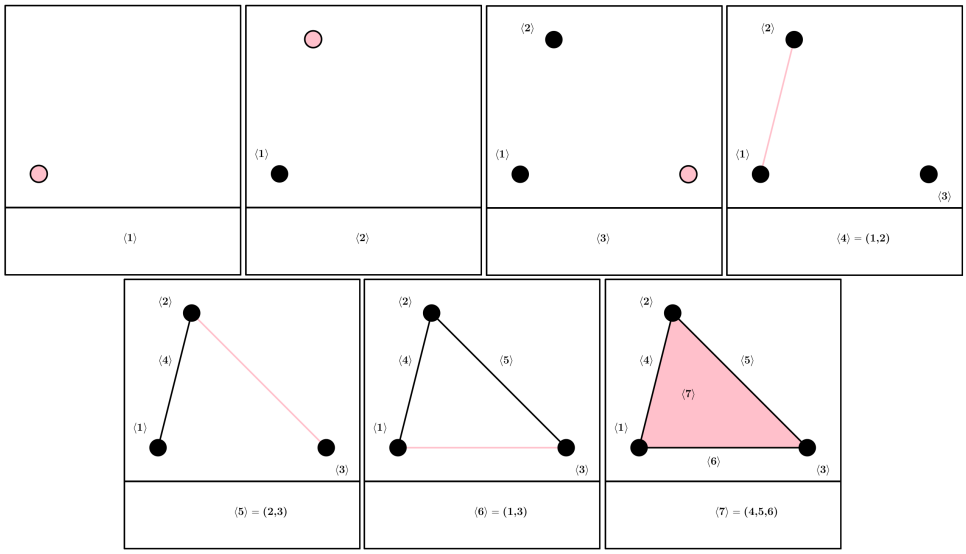


Figure 2.3 Figure illustrating the order in which the simplices of the triangle appear in the filtration.

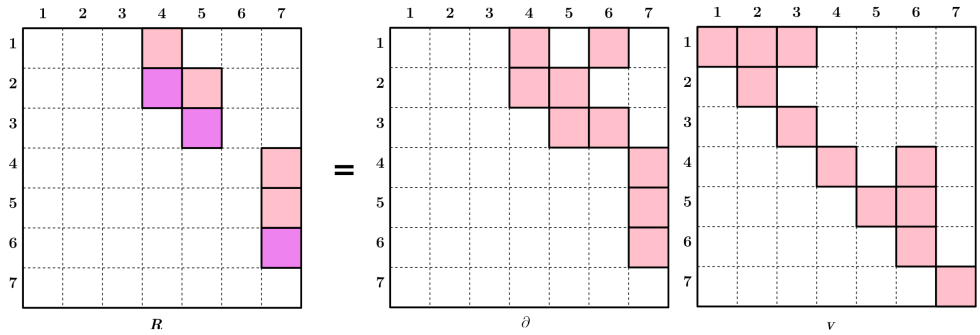


Figure 2.4 Figure illustrating reduction of the boundary matrix. R is the reduced matrix, ∂ is the original boundary matrix and V is the matrix whose column $j$encode the columns of ∂ that add up to give the column j of R. The shaded entries in the matrices denote 1. All other entries are zero.

Persistence diagrams

It is easy to read off the persistence diagrams from the reduced matrix R. In Figure 2.4, the matrix R is the reduced matrix corresponding to the persistence homology computation of the filtration of a triangle. The shaded entries in this matrix have a value 1. Moreover, the entries in a deeper shade of pink denote the lowest row of a column whose entry is 1. The lowest 1's indicate the birth-death persistence pair. In this example, the lowest 1 entry indices correspond to the set

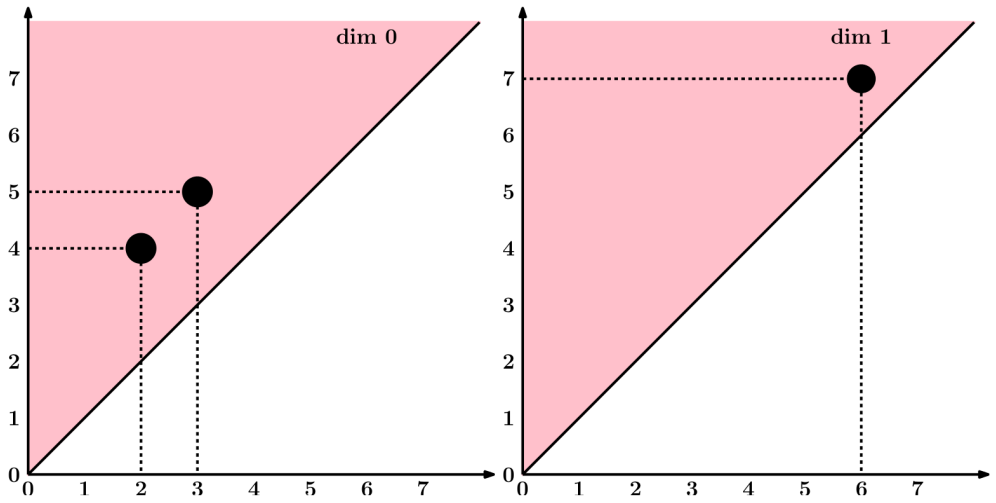


Figure 2.5 Persistence diagrams corresponding to the birth-death pairs in the filtration of a triangle. Left panel presents the 0-dimensional persistence diagram, corresponding to birth-death or merger events of isolated objects. Right panel presents the 1-dimensional persistence diagram, corresponding to birth-death events of loops.

$(i, j) = \{(2,4), (3,5)\&(6,7)\}$. The first entry in the pair is the index of the simplex that gives birth to a topological hole. The second entry is the index of the simplex that kills that particular topological hole. One can verify that the indices of these pairs indeed correspond to the birth-death pairs, as enumerated in Section 2.3.3. Figure 2.5 presents the information of birth-death pairs in the filtration of a triangle in the form of persistence diagrams.

2.3.4 Points of caution.

The methods employed in this paper are perhaps on the more sophisticated end of the spectrum of cosmic web analyzes. It is therefore important to make sure that each step is rational and reliable, and the results are not contaminated by side-effects. There are indeed a few subtleties we need to keep in mind, and we list them here to avoid possible pit-falls.

- PERIODIC TILING. Instead of the 3-dimensional Euclidean space as a model of the Universe, we use the 3-torus, which has non-trivial homology, with Betti numbers $\beta_0 = 1$, $\beta_1 = 3$, $\beta_2 = 3$, and $\beta_3 = 1$. These numbers interfere with our statistical analysis of the topology of superlevel sets, but they are barely noticeable in the midst of usually thousands for ranks we observe.

- DENSITY FIELD ESTIMATION. Among the many possible density field estimators, we rely mostly on the DTFE as it naturally adapts to the particle distribution. It has the side-effect of forming high density spikes above particles that are completely and tightly surrounded by others.

– SYMBOLIC PERTURBATION AND SUPERLEVEL SETS. We use the technical tools of symbolically perturbing the density values at the vertices, and retracting each superlevel set to the full subcomplex above the threshold. Both techniques simplify the computation but have otherwise no effect. In particular, they give precisely the same persistence diagrams and intensity plots.

– INTENSITY PLOTS. The averaged diagrams are meant to approximate the underlying distribution from which the persistence diagrams are sampled. We have no proof that they exist, other than the visual evidence that the diagrams for statistically similar particle distributions appear similar. We draw these plots by counting points within each square of a 100-by-100 grid, which implies that small shifts of the grid would give (slightly) different plots.

– PERTURBATIONS AND STABILITY. Recalling the Stability Theorem for persistence diagrams (Cohen-Steiner et al. 2007), we note an ε-perturbation of the density function can lead to the addition or removal of points at distance at most ε from the horizontal axis. As a consequence, the intensity plots may change an arbitrary amount near the horizontal axis, but not at a distance larger than ε.

2.4 Random Topology

Random processes play a crucial role in many aspects of life. In this paper, the analysis of random data provides a baseline for comparison, training the eye to pay attention to features that are not accidental, caused by inevitable random configurations in the data. We create this baseline by picking particles in space uniformly at random.

2.4.1 Poisson process.

Recall that our model of the Universe is the 3-dimensional cube with opposite faces glued to each other to create a periodic tiling of space. We call this the 3-*torus model*, denoting it by $\mathbb{X}$. We choose the length unit such that each edge is $200h^{-1}$Mpc long. Within this cube, we pick $n = 500,000$ particles in a *Poisson process*[1]. For practical purposes, the particles are thus chosen from a *uniform distribution* over the 3-torus. This forms a reasonable approximation of a Poisson process.

2.4.2 Graphs of Betti numbers.

To get a feeling for the DTF estimator of the particle sample, we compute the Betti numbers of the superlevel sets. Writing $\varrho : \mathbb{X} \to \mathbb{R}$ for the estimated density function, we plot the p-th Betti number of $\varrho^{-1}[\nu, \infty)$ as a function of ν, for $p = 0, 1, 2$. Drawing ν decreasing from left to right, we superimpose the graphs of the Betti number functions for ease of comparison; see Figure 2.6. We observe that the graph of β_0 peaks first, at a density threshold of $\nu \approx 0.04$. As expected, the graph of β_1 peaks

[1]The Poisson process depends on a parameter λ that determines the expected number of particles. We slightly rig the process such that the number of chosen particles is precisely the expected number.

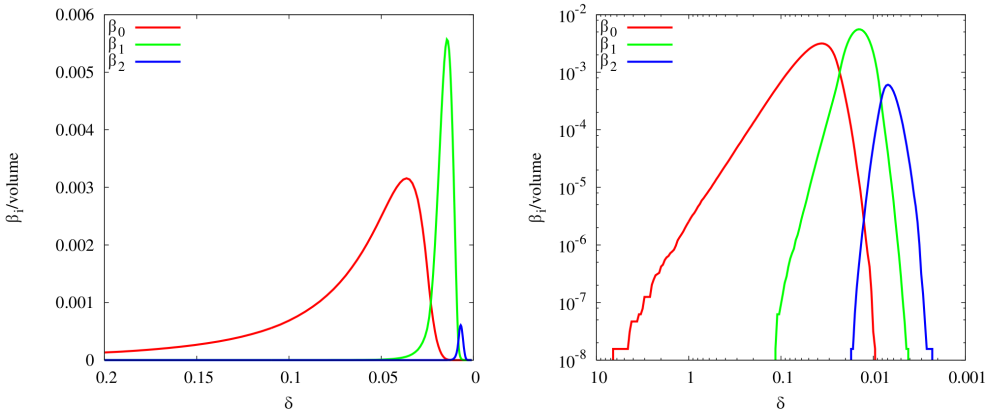


Figure 2.6 Left: The three Betti numbers of the superlevel sets of a density function on the 3-torus. The threshold, ν, decreases from left to right, and the numbers of components, tunnels, and voids increase from bottom to top. Generating 500,000 particles in a Poisson process, we get the density with the DTF estimator as explained in Appendix F.0.6. The graphs are averaged over ten realizations. Right: the same graphs in log-log scale.

second, at $\nu \approx 0.015$, and the graph of β_2 peaks last, at $\nu \approx 0.007$. This suggests that loops are formed preferably by merging clusters into filaments, as opposed to growing horns that eventually meet. Similarly, voids are formed preferably by merging clusters and filaments into walls that eventually meet to completely enclose junks of empty space. In addition to the clear order, we observe that each of the three graphs has a clean shape with a clearly defined single mode. These properties are indicative of the data following a single, well-defined distribution. Indeed, the graphs resemble that of the Poisson distribution – a single early mode and an exponentially decaying tail – which is plausible since the DTF estimator is sensitive to and preserves the local density of the particles.

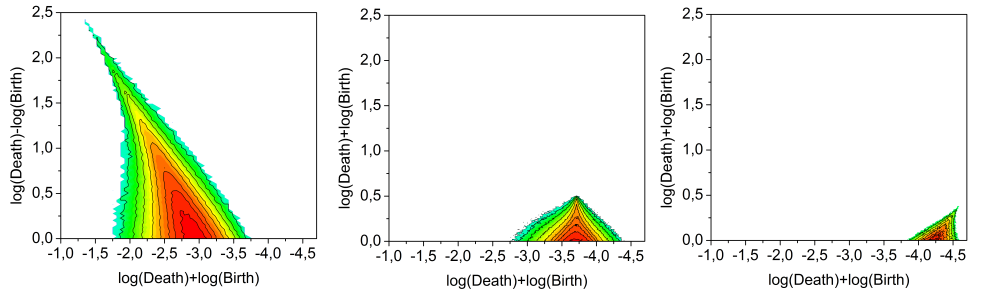


Figure 2.7 From left to right: the intensity plots of the persistence diagrams for dimensions $0, 1, 2$, averaged over ten realizations. The sum of the logarithms of birth- plus death-values decreases from left to right, while the logarithm of the persistence increases from bottom to top.

2.4.3 Averaged persistence diagrams.

As explained in Appendix B.5, persistence diagrams contain strictly more information than the graphs of the Betti numbers. Figure 2.7 shows the intensity plot of the density function, $\varrho : \mathbb{X} \to \mathbb{R}$, again in log-log scale. To compare these plots with the curves in Figure 2.6 on the right, we observe that the number of birth-death pairs, (ν_b, ν_d), with $\nu_b \geq \nu > \nu_d$ gives the Betti numbers for the superlevel set for threshold ν.[2] Since we draw the diagrams as intensity plots, we need to compare the integral over the V-shaped region anchored at the point $(\log \nu + \log \nu, 0)$ with the Betti number at $\log \nu$. When doing this, note that the horizontal axes in Figure 2.6 are labeled with values of ν, while the horizontal axes in Figure 2.7 are labeled with twice the logarithm to the base 10 of ν.

Similar to the graphs in Figure 2.6, the diagrams of β_0, β_1, β_2 are ordered along the horizontal axis. In addition, the persistence, which we see as the vertical distance from the horizontal axis, decreases from β_0 to β_1, and then again from β_1 to β_2. This is a reflection of the DTF estimator, which tends to form spikes of high density at clusters. The height of these spikes is measured by the persistence of dots in the diagram of β_0, and these spikes are visible even after taking the logarithm of the density. In contrast, the depth of voids is measured by the persistence of the dots in the diagram of β_2, which is much milder, as seen in Figure 2.7. Finally, we point out the characteristic "pointed hat" shape of the diagrams, and more specifically the sideways leaning tips for β_0 and β_2.These shapes seem related to heavily studied but difficult questions in percolation theory, and in particular to threshold phenomena, which are characteristic of this field.

2.4.4 Scaling Relations of Poisson Topology

In order to probe the scaling relations of various quantities for the Poisson distribution, we construct realizations with different mean inter-particle separation $\lambda = 0.0625, 0.125$ and 0.25. Keeping the box size same, this amounts to an increased number of particles with decreasing λ. Figure 2.9 plots the Betti numbers for realizations with different λ, where the horizontal axis (corresponding to level set value) is scaled with the variance of density, which is computed at each particle position. The β_is for different λ's have the same peak positions after scaling. Peak positions are well separated, denoting that topology is predominantly either "cluster-like", "sponge-like" or "cheese-like" at different level set values. β_0 peaks at $\nu \approx 1.8$, β_1 at $\nu \approx 0.6$ and β_2 at $\nu \approx 0.3$. The coincidence of peak-positions suggest a functional form of Betti numbers as a function of level set.

In addition to the scaling of peak positions with normalized level set values, the peak amplitudes and the location of the peak amplitude of β_i also scale with λ. This scaling is shown in the top-left and top-right panels of Figure 2.8. Peak amplitudes of β_0, β_1 and β_2 scale linearly with λ, with different slopes. β_1, the number of loops rises the sharpest with λ, with a slope of $m = 0.08902$, followed by β_0 ($m = 0.05036$) and β_2 ($m = 0.00989$). The non-normalized (with respect to variance) peak positions on the horizontal axis also scale with λ. However, the trend is not the same as the

[2] This relation may be violated by the $8 = 1+3+3+1$ essential homology classes of the 3-torus, which are not drawn in our diagrams. Their number is too small to be notices in our figures.

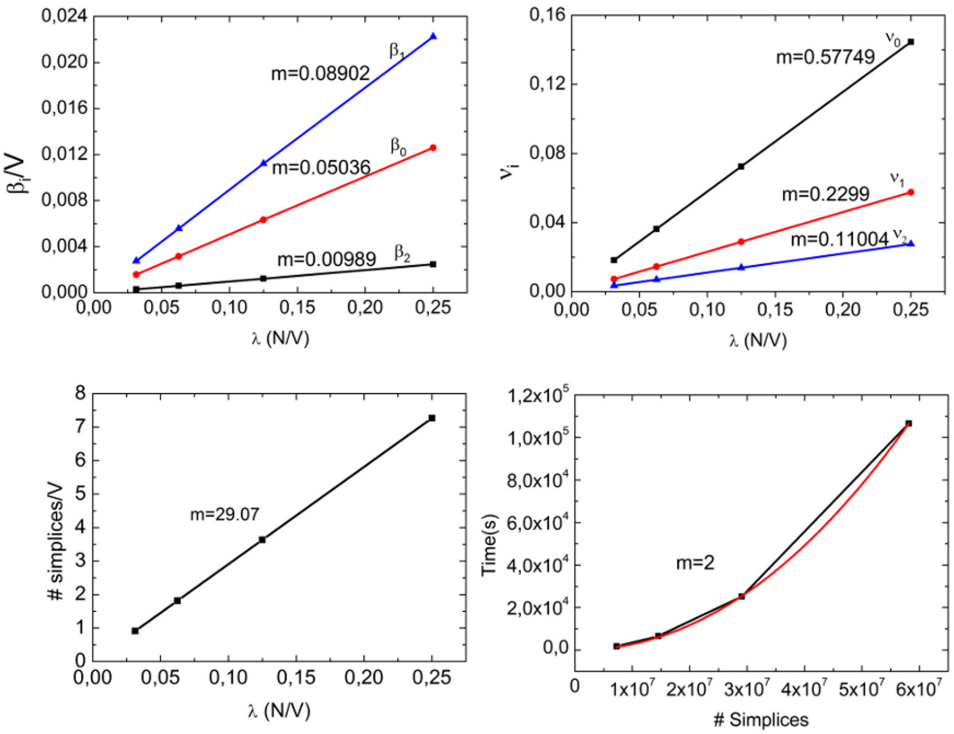


Figure 2.8 Scaling relations for different quantities for the poisson distribution. The quantities on the vertical axis (except the bottom-right panel) are per-unit-volume. Top-left panel : Scaling of peak-amplitude of β_0, β_1 and β_2 with the parameter of distribution λ. Top-right: scaling of normalized (with the standard deviation) peak-position (on the horizontal axis) with λ. Bottom-left: scaling of number of simplices with *lambda*. This can be translated to the scaling of number of simplices with the number of particles in the box. Bottom-right : scaling of time required to compute persistence with the number of simplices. The quantities on vertical axis scale linearly with quantities on horizontal axis in the top-left, top-right and bottom-left panel. The scaling in bottom-right panel has a power-law form. the slope of scaring is doted by "m" in the first three panels. In the fourth panel, m is the index of the power-law distribution.

peak amplitudes. In this domain, ν_0, the peak position for β_0 rises the sharpest with increasing λ, with a slope of $m = 0.57749$, followed by ν_1 ($m = 0.2299$) and ν_2 ($m = 0.11004$), in that order.

The number of simplices per unit volume also scales linearly with λ and has a slope of $m = 29.07$. This is presented in the bottom-left panel of Figure 2.8. Bottom-right panel of Figure 2.8, presents the scaling of time required to compute persistence for the poisson distribution with respect to the number of simplices in the triangulation. The time required to compute persistence seems to follow a power-law with respect to the number of simplices. We fit a power-law of the form $f(x) = ax^b$ where b is the index of the power-law. The fitted curve to the data points gives the value of the index $b = 2$.

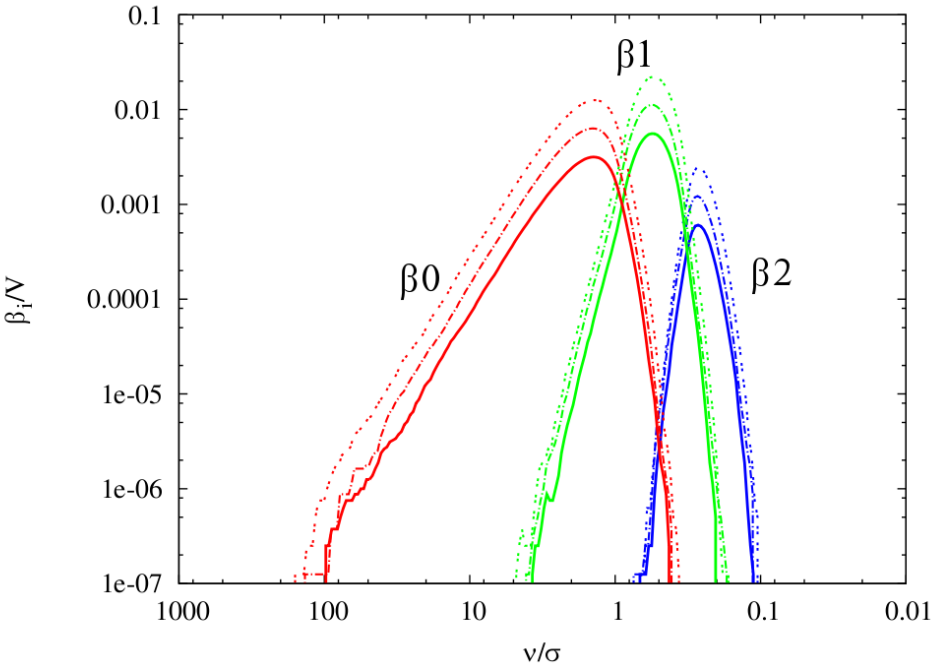


Figure 2.9 Betti numbers for poisson distribution with λ, the parameter of distribution varying. For each realization, the levelset values on the horizontal axis are normalized by the standard deviation of that particular realization. In the representation of normalized horizontal axis, the peak positions for realizations with different λ are coincident. The lowest peak-amplitude corresponds to $\lambda = 0.0625$, followed by $\lambda = 0.125\&0.25$ respectively.

2.5 Single-Scale Topology

In this section, we consider a random process that produces particle distributions near the elements of a fixed Voronoi diagram. While heuristic in nature, these distributions mimic the structural patterns observed in the Universe: the clusters, filaments, and walls in the Cosmic Web.

In these Voronoi clustering models, outlined in appendix D, a geometrically fixed Voronoi tessellation defined by a small set of nuclei is complemented with a heuristic prescription for the location of particles or model galaxies within the tessellation (van de Weygaert & Icke 1989; van de Weygaert 1991; Weygaert 2007). We distinguish two classes of Voronoi models: the *pure Voronoi element models* and the *Voronoi evolution models*. Both are obtained by moving an initially random distribution of N particles toward the faces, lines, and nodes of the Voronoi tessellation. They do this by a heuristic and user-specified mixture of projections onto the various geometric components of the tessellation. The Voronoi evolution models accomplish this via a gradual motion of the galaxies from their initial, random locations towards the boundaries of the cells.

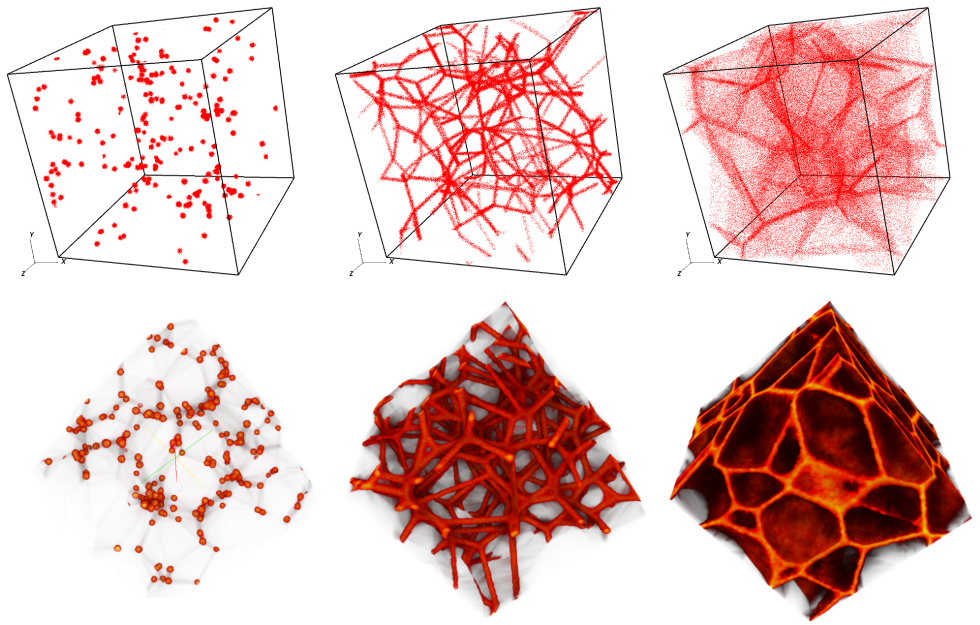


Figure 2.10 Top row, from left to right: particle distribution in the three pure Voronoi element models corresponding to clusters, filaments, and walls. Each data set consists of 262,144 particles inside a periodic box of side length $200h^{-1}$Mpc. Bottom row, from left to right : density rendering of the same.

2.5.1 Pure Voronoi element models.

Recall that a Voronoi diagram in space has four types of elements: vertices, edges, faces, and cells. Constructing and fixing a diagram for only 32 nuclei within a periodic box with sides of length $200h^{-1}$Mpc, we consider three random processes that generate particles near the vertices, edges, and faces. With each realization, we get 262,144 particles distributed uniformly along and with a Gaussian spread of $1h^{-1}$Mpc around the elements of the Voronoi skeleton; see Figure 2.10 and appendix D for a detailed description of the models. The first process generates the particles in *clusters* around the vertices, the second forms *filaments* along the edges, and the third creates *walls* following the faces. Since each process focuses on the elements of a single dimension, we call the resulting distributions *pure Voronoi element models*.

2.5.2 Graphs of Betti numbers.

We begin our analysis by looking at the Betti numbers of the superlevel sets of the estimated density field. Figure 2.11 shows the numbers as functions of the threshold. All results are averaged over eight realizations. The number of particles being the same in all three models, the average matter density in the clusters is higher than

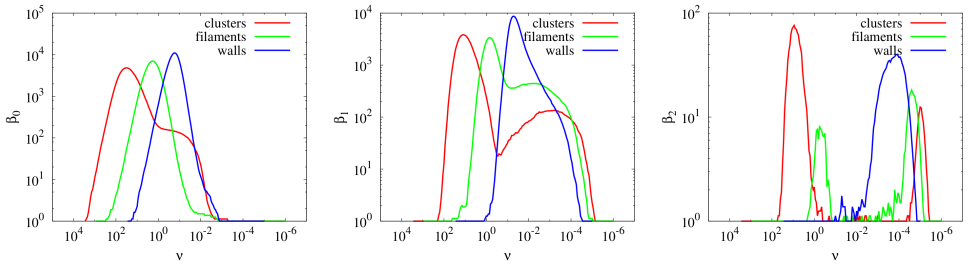


Figure 2.11 The Betti numbers of the superlevel sets of the density function for pure Voronoi element models as functions of the threshold. From left to right: β_0 β_1, β_2.

along the filaments, which in turn is higher than inside the walls. This is reflected by the graphs of β_0, in which the density threshold of the maximum is highest for clusters, between the extremes for filaments, and lowest for walls. The value at the maximum (the number of components) follows a reverse trend.

Note the prominent shoulder in the graph of β_0 for clusters, which we do not see in the graphs for filaments and voids. The shoulder is a reflection of the merging process, which first consolidates the particles into clusters and second merges the clusters into one connected whole. We thus observe a transition from *intra-cluster* to *inter-cluster* merging, with the parameters of the shoulder identifying the density values at which this transition happens. In the filament and wall models, we have a single connected component as soon as all filaments and walls have been consolidated, which explains the absence of shoulders. Nevertheless, we observe a transition from a focus on intra- to inter-structural connectivity as a function of the density threshold. Indeed, the graph for β_1 has a shoulder, both for clusters and for filaments, and the explanation is similar.

Continuing the trend, the graph for β_2 has two clear modes for clusters and filaments, and a hint of two modes for voids. A comparison with the intensity plots shows that this hint is a fluke, and while the separation into two populations of voids is real, it is not visible in the graph. More about this shortly. Returning to the graphs of β_2, we note that the left modes reflect the consolidation of the particles sampling the Voronoi elements, and the second modes reflect the filling up of the global, inter-structural voids. We see that the ordering of the left modes from clusters to filaments to walls is reversed for the right modes, remembering that β_2 for walls does not distinguish between the two populations and combines the left and right modes into one. The reversal of order makes geometric sense, since we are talking about the same voids in all three models, but these voids are shallower and appear at lower density values for clusters than for filaments, and more so for walls.

2.5.3 Averaged persistence diagrams.

The intensity plots for the pure Voronoi element models display features the graphs of the Betti numbers fail to capture, primarily because the plots distinguish between significant and insignificant features. For example, each realization of the filament

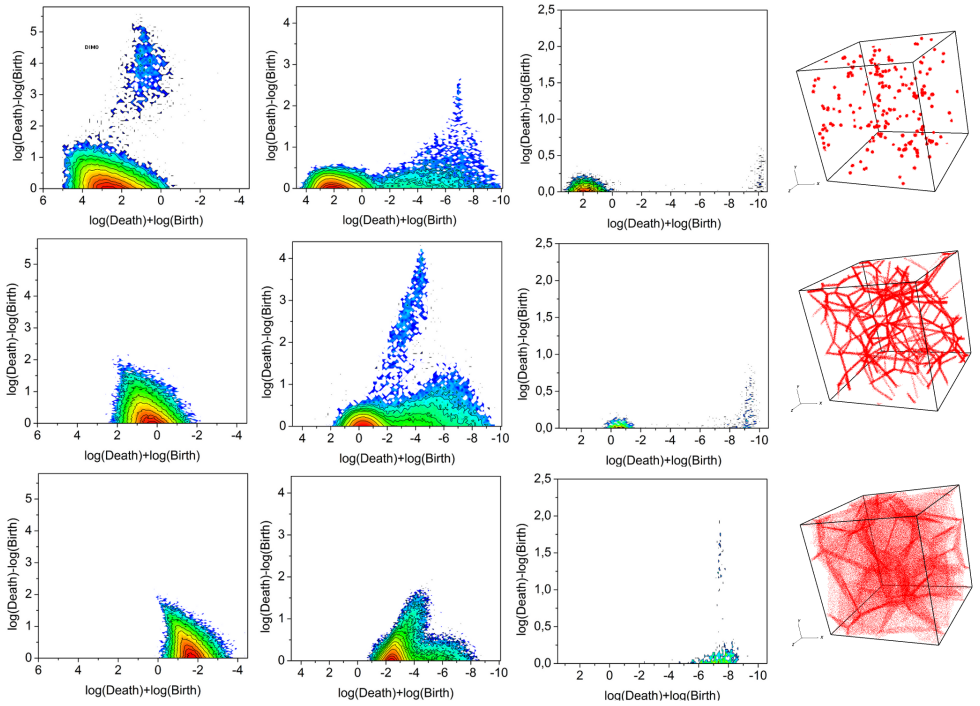


Figure 2.12 The averaged persistence diagrams of the density functions for pure Voronoi element models. From top to bottom, we show the intensity for clusters, filaments, walls, and from left to right for classes of dimension $0, 1, 2$.

model has a large number of tiny loops inside the filaments, but also a smaller number of larger loops that are carried by the filaments themselves. The 1-st averaged persistence diagram distinguishes between these two populations.

More generally, Figure 2.12 shows the intensity plots of all diagrams for all pure Voronoi element models: from top to bottom for clusters, filaments, voids, and from left to right for β_0, β_1, β_2. To a first degree of approximation, all diagrams contain a red and green high-intensity region and a blue low-intensity region. For the six diagrams in the upper-right triangle of the 3-by-3 array, the second region forms a *island*, by which we mean a hill that is completely surrounded by a ring of zero intensity. As before, the high-intensity regions reflect the intra-structural consolidation, while the low-intensity regions consist of points that represent large topological structures each carried by several clusters, filaments, or walls. For components, the two populations are clearly separated in the upper-left diagram for clusters.

Similar to the graphs, we see no separation into the two populations of components in the diagrams for filaments and walls. For loops, the two populations are most clearly separated in the center diagram of Figure 2.12, which plots the intensity for filaments. The two populations of loops are less clearly separated in the top diagram for clusters, and not at all separated in the bottom diagram for walls. Nev-

ertheless, that plot has a tongue suggesting a population of loops emigrating from the bulk. The geometric interpretation of this phenomenon is that the walls meet in filaments, which are therefore more densely sampled, so that global loops can form before the walls are completely filled.

For voids, the separation into two populations is clearly visible in all three diagrams; see the third column in Figure 2.12. Most noteworthy is the separation in the bottom diagram, in which the two populations have roughly the same mean age but very different persistence. Such populations cannot be separated by V-shapes, which is the reason the function of Betti numbers is oblivious to this difference.

2.6 Multi-Scale Topology

One of the major features of the matter distribution at large scales is the presence of a hierarchy of substructures, with a large dynamic range in density and spatial scale. As a result, we see a multi-scale distribution, with interesting features at every scale.

2.6.1 the Soneira-Peebles model.

Soneira-Peebles is a random process with adjustable parameters that generates a fractal distribution of particles (Soneira & Peebles 1978). Both the two-point correlation function and the fractal dimension of these particle sets are well understood analytically; see Appendix E. The parameters can be chosen such that the correlation function of the particle distribution mimics that of the galaxies in the sky. It is used to explain the clustering statistics of the galaxy distribution, taking into account the fact that they display strong self-similarity. The placement of the particles is controlled by three parameters, each responsible for tuning a different aspect of the hierarchy:

η**:** the *height*, equal to the number of levels minus 1;

ζ**:** the *concentration*, equal to the ratio between consecutive radii;

ψ**:** the *branching factor*, equal to the number of children.

We start the construction with a unit sphere at level 0, inside which we place the centers of ψ level-1 spheres, each with radius $1/\zeta$ at random positions. The next iteration places the centers of ψ level-2 spheres with radius $1/\zeta^2$ inside each level-1 sphere. We continue the process until we reach level η, with a total of ψ^η spheres of radius $1/\zeta^\eta$. Finally, we pick a particle at the center of each level-η sphere. Figure 2.13 shows three sample distributions with fixed height and branching factor, but with varying concentration.

2.6.2 Graphs of Betti numbers.

We study particle distributions generated with height $\eta = 6$, branching factor $\psi = 9$, and three different concentrations, $\zeta = 5.0, 7.0, 9.0$. For each parameter triplet, we average the results over eight realizations. Figure 2.14 shows the Betti numbers as functions of the threshold defining the superlevel set of the density functions defined

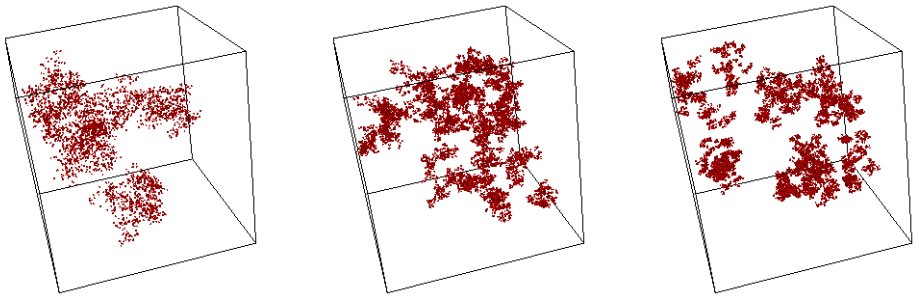


Figure 2.13 Particle distributions generated with the Soneira-Peebles process. Fixing the height to $\eta = 6$ and the branching factor to $\psi = 9$, we vary the concentration from left to right as $\zeta = 5.0, 7.0, 9.0$. there are 6^9 particles in each data set. Due to the high concentration factor, the number of particles may seem smaller than that, but it is not. Zooming into a particular region shows similar structure at higher levels of hierarchy. Density rendering of the distribution is not feasible due to high concentration.

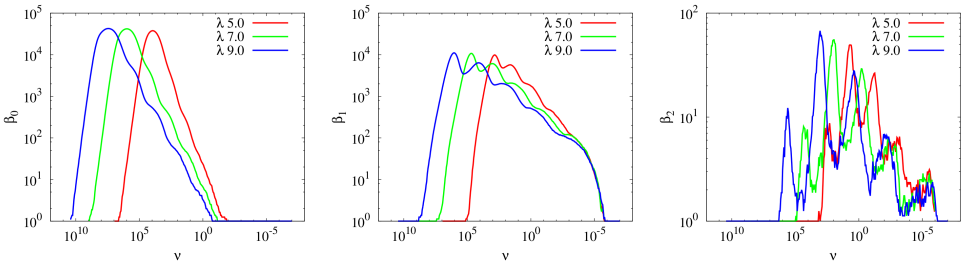


Figure 2.14 From left to right: the 0-th, 1-st, 2-nd Betti numbers of the superlevel sets of the density function for the Soneira-Peebles particle distributions plotted on a logarithmic scale. Fixing the height to $\eta = 5$ and the branching factor to $\psi = 9$, we vary the concentration as $\zeta = 5.0, 7.0, 9.0$.

by the particle distributions. Evidence of modularity[3] is present in the curves for all chosen values of ζ. For β_0, it manifests itself as ripples on the right side of the mode, when the number of components decreases after reaching a maximum. For β_1 and β_2, the evidence can be seen in the number of modes. Higher concentration results in a more clearly defined modular distribution. Indeed, the number of distinct ripples in the graphs for β_0 is the largest for $\zeta = 9.0$, while they are barely visible for $\zeta = 5.0$.

The peak amplitude for β_0 is the same for all three distributions. The reason may be trivial, namely the fact that η and ψ are the same for all three experiments, implying that all data sets contain the same number of particles, namely ψ^η. However, the peaks occur at different density thresholds, reflecting the varying local density of the distributions generated for different concentrations. Indeed, more concentrated particle distributions have higher density peaks, and as a result we see the mode at

[3]The term "modularity" is used for particle distributions with distinguishable levels in the hierarchy. A *modular* distribution is hierarchical in nature.

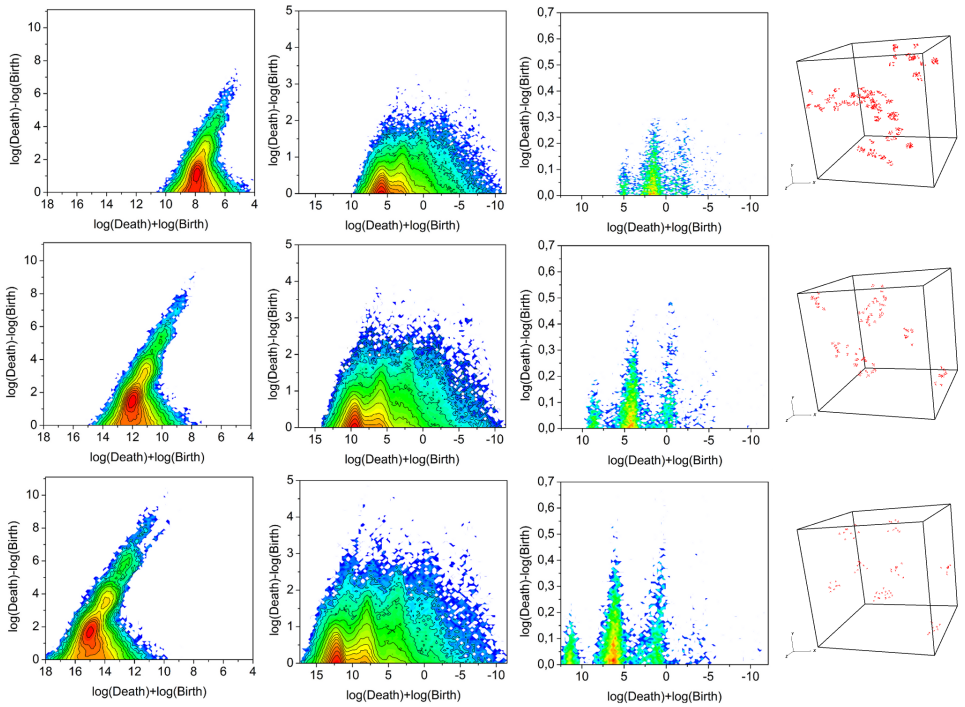


Figure 2.15 From left to right: the 0-, 1-, 2-dimensional persistence diagrams of the density functions obtained from the Soneira-Peebles particle distributions. Fixing the height to $\eta = 5$ and the branching factor to $\psi = 9$, we vary the concentration from top to bottom as $\zeta = 5.0, 7.0, 9.0$.

higher thresholds. We observe the same trend in the curves for β_1, and even for β_2, although the latter curves a much rougher, reflecting overall smaller numbers and more noise. The number of levels in the hierarchy is reflected in the number of ripples, which is most clear for the graph of β_1. We see five distinct ripples, while the number of levels in the distribution is six. It seems that the lowest level has too few components to be visible in the graphs. While the graphs of β_2 are noisy, they also exhibit a similar ripple structure.

2.6.3 Averaged persistence diagrams.

The intensity plots of the particle distributions described above are shown in Figure 2.15, for $\zeta = 5.0, 7.0, 9.0$ from top to bottom, and for dimension 0, 1, 2 from left to right. The features in the diagrams show a clear transition as a function of the concentration, with evidence of modularity present in all diagrams. In particular, we notice *hills* in the intensity, which we define as the neighborhood of a local maximum away from the horizontal axis. Note that these are different from *tongues* in the intensity plots, which are regions right below the local persistence maxima.

Hills seem rather unusual features as the intensity usually decreases monoton-

ically from bottom to top. For the 0-dimensional diagrams, we notice an increase in the number of hills when we increase the concentration: there is a single hill for $\zeta = 5.0$, we see the hint of a second hill for $\zeta = 7.0$, and there are three clear hills for $\zeta = 9.0$. In words, we get progressively more evidence for modularity as the concentration increases, which is hardly surprising. Interestingly, the hills come in sequence, from bottom to top, so that later hills represent birth-death pairs of higher persistence. Furthermore, the intensity of the hills decreases from bottom to top. This makes sense since lower levels in the construction contain fewer clusters with lower persistence. Indeed, the highest level in the hierarchy generates the densest regions with the largest number of particles. Physically this means that many tiny clusters form at high density thresholds. These clusters are short lived, and as we go down from the highest level, a large number of tiny clusters merge together to form fewer but larger clusters. These larger clusters are of higher persistence and correspond to the low-intensity, high-persistence hills in the diagrams. The bias of the higher persistence hills towards the lower density values, is interesting, as it counters the higher density leaning pointy hat shape we see for the uniformly distributed particles; see Figure 2.7.

Progressively better defined modularity as a function of increased concentration is also evident in the 1-dimensional intensity plots. Here, we see tongues that correspond to the hills in the 0-dimensional plots. Larger concentration corresponds to smaller filling rate, which results in bigger patches of empty space. This is reflected in the 2-dimensional intensity plots, which record the information for the voids or empty regions: we see three or perhaps four grainy tongues, which are fuzzy for $\zeta = 5.0$, and progressively better defined for $\zeta = 7.0$ and 9.0.

2.7 Dynamic Topology

In this section, we consider particle distributions that change configuration over time, similar to the matter in the Cosmos. Under the influence of gravity, the relatively uniform distribution at early epochs accumulates in the potential wells, evolving into galaxies arranged in clusters, filaments and walls.

2.7.1 Voronoi evolution models.

Starting with a random distribution of particles over the entire volume, *Voronoi evolution* generates a time-series of particle distributions driven by slow drifts from higher- to lower-dimensional elements of an underlying Voronoi diagram. Their construction is described in some detail in appendix D.

We have sampled this time-series at three moments in time, called *stages*, and we show the results for these, emphasizing the continuous change that becomes visible by comparing the graphs and diagrams. To parametrize the stages, we keep track of the percentage of particles that lie in the interior of cells, faces, edges, and vertices of the Voronoi diagram; see Table 2.1. Stage 1 is the least evolved particle distribution, With the highest percentage of particles in cells, while Stage 3 is the most evolved distribution, with the highest percentage at and around the vertices. Figure 2.16 shows the three stages as point clouds, going from left to right in the evolution.

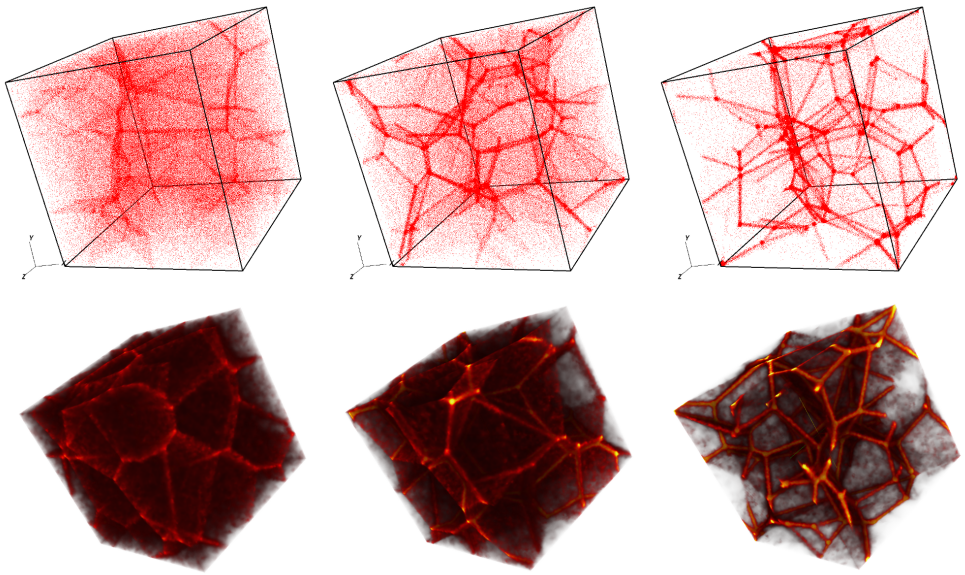


Figure 2.16 Snap-shots in the Voronoi evolution time-series. Top row, from left to right: particle distribution at the least, medium, most evolved stage. Bottom row, from left to right: volume rendering of the same.

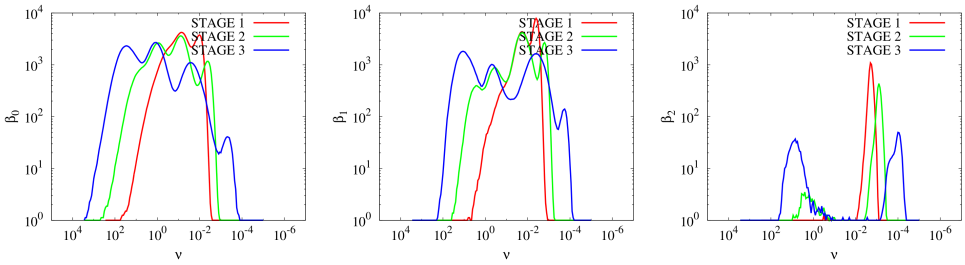


Figure 2.17 The graphs of the Betti numbers computed for the suoerlevel sets of the density function of evolving particle distributions. From left to right: $\beta_0, \beta_1, \beta_2$ at different stages of the evolution. Stages 1, 2, 3 progress from least, to medium, to most evolved.

	cell	wall	filament	cluster
Stage 1	49.93%	38.52%	10.46%	1.08%
Stage 2	5.03%	23.50%	41.26%	30.22%
Stage 3	2.00%	14.72%	39.81%	43.47%

Table 2.1 The relative abundance of particles in each structural element throughout the course of evolution. Stage 1 is the least evolved, with almost half the particles residing in cells, while Stage 3 is the most evolved, with almost half the particles residing in clusters.

2.7.2 Graphs of Betti numbers

We show the Betti numbers as functions of the threshold defining the superlevel set in Figure 2.17. The graphs are significantly different from the ones we see for the single-scale Voronoi models in Figure 2.11. The graphs for β_0 show a gradual transition from two to four peaks. The four peaks in Stage 3 reflect the fact that we have a non-trivial number of particles populating each of the four morphological features (clusters, filaments, walls, and the space in between) so that each population contributes its own peak to the graph. As before, the contributions are ordered from left to right as the clusters are densest and merge first, and so on. In contrast to Stage 3, Stage 1 has most particles near the walls and in the space between them, so that there are only two significant contributions to the graph.

A similar trend is also seen in the graphs for β_1. The particle distribution gets progressively more segregated into the morphological features, each with its own density, which explains the clear four peaks we see for Stage 3. The signal we get from β_2 is different while consistent with our explanation. We see one peak at Stage 1 and two peaks each at Stages 2 and 3. As before, the difference is between intra- and inter-structural consolidation, and the second peak barely exists in Stage 1, at which time a large fraction of the particles populates the space between the walls.

2.7.3 Averaged persistence diagrams

The evolution of the particle distribution is well visible in the averaged persistence diagrams, which we show separated for the three stages and the different dimensions in Figure 2.18. Each intensity plot is obtained by averaging eight realizations. While the evolution flows from top to bottom, we show the results for the components, loops, and voids from left to right.

Recall that Stage 1 is dominated by particles distribution near the walls and in the space between the walls. Corresponding to the two peaks of the graph for β_0, we see two tongues in the upper-left intensity plot, which shows the averaged diagram for the components. Note that the tongue with higher intensity is on the right hand side, where the mean age is smaller. Indeed, the density in the space between the walls is smaller while the population there is larger. Two things happen when we go from Stage 1 to Stage 3: the number of tongues increases to four, and the order of the tongues by intensity is reversed. Similar to the graphs of the Betti numbers, we attribute the four tongues at Stage 3 to a clean segregation of the particles into four morphological elements. The change in order is of course due to the trend to put larger populations of particles into lower-dimensional elements. We point out that the two phenomena are related to each other. The percentage of particles in a morphological component dictates its average density, which, in turn, drives the segregation.

Note also the formation of a low-intensity island in the intensity plots, which breaks from the bulk and migrates towards high persistence values as the model evolves. We see this phenomenon in all three dimensions. The underlying reason is that the cells deplete of particles during the evolution, and the created empty space favors the appearance of inter-structural consolidation – a manifestation of the structure of the underlying Voronoi skeleton itself – which is represented by the islands.

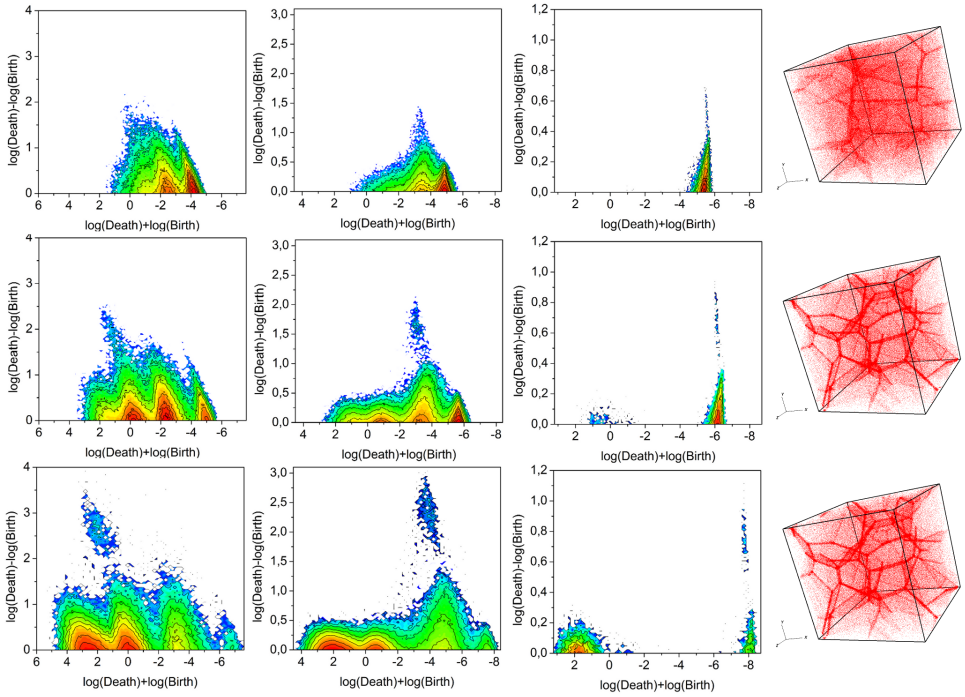


Figure 2.18 The averaged persistence diagrams of the density function for the Voronoi evolution models. From top to bottom, we show the intensity plots for least, medium, most evolved stages, and from left to right for classes of dimension $0, 1, 2$.

2.8 Persistence Based Noise Estimation.

The density field arising out of the matter distribution in the Universe is a continuous smooth field. However, the current methods employed by us to study the Universe are based on discrete techniques, almost without exception, both in cosmological simulations and observations. Simulations are n-body experiments that use discrete particles to model the Universe, while closer home observationally, we rely on galaxies as the luminous tracers of the baryonic matter. These galaxies are assumed to be embedded in dark matter, and hence assumed to trace the dark matter distribution as well. Because of this discrete sampling of density field in simulations and observations alike, there is an inherent noise in the measurement of quantities.

We expound on a method that employs the concept of persistence to set a benchmark for noise, and thereby segregate noisy structures from significant ones. We use the single component Voronoi models to test the credibility of noise-estimation. For these models, we can estimate the volume occupied by the structural elements in the box. We also know the number of sample points used to construct these models, *a-priori*. Though in our construction of the models, the points are sampled around the voronoi elements with a finite thickness, and the sampling density follows a Gaus-

Model	# particles	Threshold	Volume Occupancy	Edge Length of Poisson Box
Cluster	262,144	1	78,914.6	43
Filament	262,144	1	371449.5	72
Wall	262,144	1	1525370.2	115

Table 2.2 Parameters for determining the poisson distribution for noise estimation for the single component voronoi models. Volume is estimated by extrapolating density to a regular grid cell using volume-weighted DTFE, and adding the volumes with grid cells with density threshold above 1. Column 1 denotes the model investigated, column 2 the corresponding number of particles. Column 4 is the volume occupied by the structural elements. Column 2 and 4 are used to estimate column 5, the edge-length of the Poisson-distributed box with the same number of particles and volume occupancy as the model in question.

sian profile, we assume that they are uniformly sampled in the given thickened element. We will demonstrate shortly that this assumption has no significant negative repercussions on the detection of significant structures.

The recipe for noise estimation is the following : for a given number of particles and volume occupancy, we construct a box with same the same volume and number of particles, where the particles are poisson distributed. This is in essence equivalent to constructing a poisson distribution with the same mean inter-particle separation as inside the structural elements of the voronoi model. The parameters for construction are presented in Table 2.2. The Poisson distribution has a characteristic persistence diagram for a given mean inter-particle separation. We set the maximum persistence, or the height of the persistence diagram for this poisson distribution as the persistence threshold of the noisy structures for the respective models in question.

Figure 2.19 presents the persistence diagrams for $0-, 1-$ and $2-$ dimensions (in red) superposed with the respective diagrams for the poisson distribution (in blue). The black line denotes the persistence threshold for noise as determined from the poisson distribution. For the $0-$ dimensional diagrams, we notice that there are significant number of dots above the persistence threshold for clusters, while this is not the case for filaments and walls. This is because for the filament and the wall models, by the time we cross the persistence threshold, the structure has consolidated into a single connected component. For the cluster model, however we still have many isolated objects, which are the clusters themselves. In $1-$ dimension, there is a formation of loops which have higher persistence than the noise threshold for all the three models. For the cluster model, it arises out of the inter-connectivity of individual clusters as we decrease the superlevel set. For filaments and walls this is a manifestation of the loops formed by the particle distribution, a significant number of which are contributed from the loops formed by the edges of the underlying voronoi skeleton. In $2-$ dimensions, we notice that there are very few persistent dots above the noise threshold. This is because the *mass-weighted* DTFE heavily under-samples the void regions (as opposed to cluster, filament and wall regions, where the sampling density is high). To get a better estimate of structures in the voids, we propose the use of *volume-weighted* DTFE. We also notice that the dots in the main island for the voronoi models have a wider spread along the horizontal axis compared to the pois-

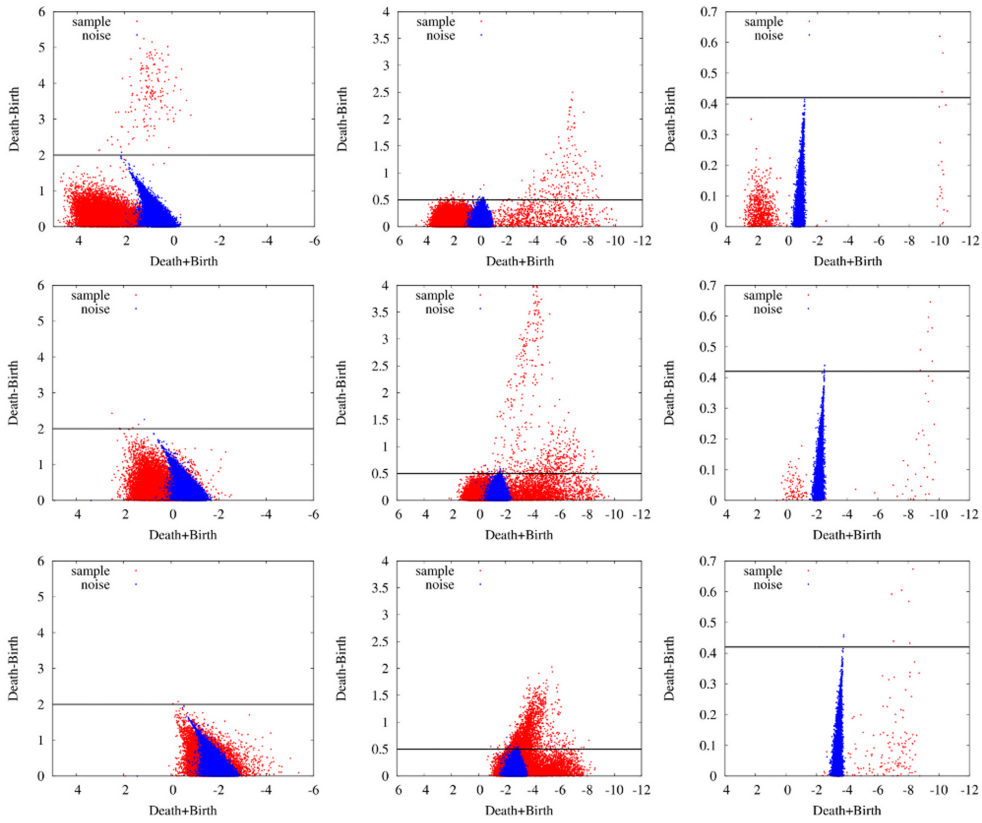


Figure 2.19 Persistence diagram for single component Voronoi models (in red) superposed with the corresponding diagrams of appropriately constructed poisson distribution (in blue). Top to bottom: diagrams for clusters, filaments and walls respectively. Left to right: $0-$, $1-$ and $2-$dimensional diagrams. We set a the benchmark for noise threshold at the maximum persistence of the diagram for poisson distribution, which is at approximately $\tau = 0.5$ for all the models.

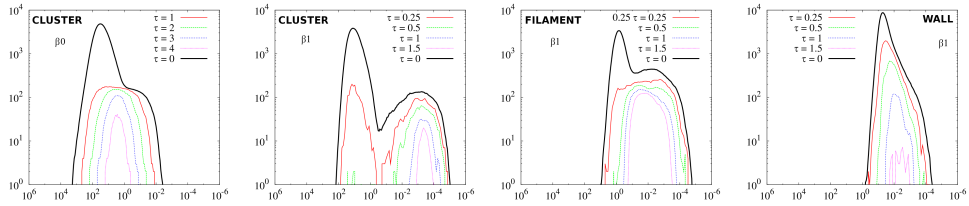


Figure 2.20 Persistent Betti numbers in 1 dimension, as derived from the Poisson-benchmarked diagrams in Figure 2.20. Left to right: clusters, filaments and walls respectively. The curves corresponding to $\tau = 0.5$ count the loops which are exactly above the poisson threshold. The curves corresponding to $\tau = 0$ count all the loops which are above zero-persistence threshold – in other words, the original Betti numbers.

son distribution. This is due to the fact that that the dots are not Poisson-sampled around the elements, but rather follow a Gaussian profile. However, this seems to have little effect on the maximum persistence of noisy structures, which is what we predicted above.

2.8.1 Persistent Betti Numbers.

Using the persistence diagrams, we can correct for noise in the Betti number graphs as well. The *persistent Betti numbers* for a level ν get contributions from all the cycles which are born before ν, die after ν and have a persistence larger than a given threshold.

Figure 2.20 plots the persistent Betti number curves for panels that have significant number of cycles above the persistence threshold in Figure 2.19: $0-$dimensional structures (isolated objects) for clusters, and $1-$dimensional structures (loops) for clusters, filaments and walls. The curve in black denotes the regular Betti number curve, counting all the structures at a given threshold, without employing persistence for filtering them. The rest of the curves are for cycles above different persistence threshold as determined from the persistence diagram in Figure 2.19.

Denoting τ as persistence, for the 0-dimensional diagrams in cluster model we fix the threshold at $\tau = 2$. For the 1-dimensional diagrams for clusters, filaments and walls, we fix the threshold at $\tau = 0.5$. We notice a complete suppression of the prominent peaks at high density thresholds for the first three panels in Figure 2.20. As we had explained in 2.5, these peaks correspond to noisy structures, and are corrected for, once we take into account the persistence threshold for noise. For clusters, the persistent Betti numbers for $\tau = 2$ in 0-dimension start appearing at density value of around 0.01 and peak at around 0.001 before they start dropping for yet lower density thresholds. At their maximum, we count above a 100 clusters above noise threshold. For the $1-$ dimensional persistent Betti numbers, setting the persistence threshold at $\tau = 0.5$ as determined from the diagrams, we count about a 100 loops for clusters and filaments. The number is significantly higher for walls.

2.9 Density estimators: comparison

Within the computational pipeline outlined in Appendix B, we have encountered the issue of density field representation. This involves the choice for both the density field estimator as well as the field sampling. For a full appraisal of the potential of persistent homology in the analysis of the cosmic web, we need to understand the sensitivity of persistence diagrams and Betti number curves to the density estimator and field sampling.

The most widely used choice of density estimation is the kernel density estimator. It has specifically been used to compute persistent homology of superlevel sets of manifolds, in the context of defining confidence intervals for delineating noise from real signal (Fasy et al. 2013). Bobrowski et al. (2014) apply the kernel density estimator to introduce a consistent estimator for homology, that involves inferring the homology structure of manifolds from noisy data. Kernel density estimates have also been used to study the possibility of inferring the geometric inference of a point

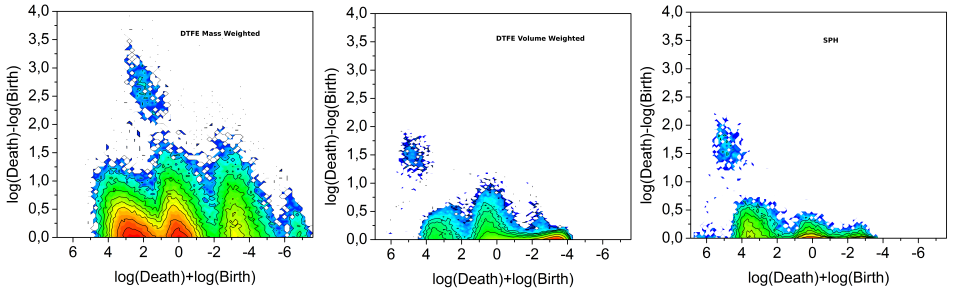


Figure 2.21 The averaged 0-dimensional persistence diagrams for the superlevel sets of the density functions estimated from point sets generated according to the Voronoi evolution model. From left to right: the intensity plots for the density estimated according to the mass-weighted DTFE, the volume-weighted DTFE, and the SPH method.

cloud (Chazal et al. 2011) by using the kernel distance (Phillips et al. 2013), instead of distance to a measure (Chazal et al. 2009; Chazal et al. 2011).

In this chapter, we use the DTFE technique to estimate density. It is an important exercise to investigate the effects of the choice of different density estimators on the computational results presented in this chapter. In view of this, in this section, we compare the resulting persistence diagrams for three different recipes of density estimation. The first procedure involves field estimation using the DTFE technique on the raw input point cloud. The second method produces an image on regularly sampled grid points. The third procedure that we test is the standard kernel density estimator. A brief account of the technical details o the different density estimators is provided in Appendix F.

The tests are based on eight model realizations, each of them corresponding to the most developed stage of the Voronoi evolution model (see table 2.1). It suffices to pick up 0-dimensional intensity plots as a representative example.

2.9.1 Density field sampling strategies

The first comparison is between a computational pipeline built on the raw DTFE density field sampling strategy and that based on DTFE image sampling of the density field on a regular grid. The raw DTFE sampling strategy operates on the basis of the DTFE Delaunay triangulation and the density field estimates at its vertices. The DTFE image sampling strategy samples the density fields on a regular grid, and subsequently processes the 3-D image for extracting its persistent topology.

The first two frames of figure 2.21 show the resulting intensity plots for the raw DTFE sampling strategy (lefthand frame) and the image DTFE sampling (central frame). The two plots share several important features. Each plot consists of two regions: a low-intensity, high-persistence island and a high-intensity, low-persistence mainland that is structured into several tongues. The island reflects the underlying Voronoi skeleton in which high-density clusters persist over a large range of scales.

The range is more pronounced for the raw DTFE sampling than for the image

DTFE sampling. It is reassuring that the basic information conveyed by the three plots is qualitatively the same and thus independent of the particular density estimator used to construct the function under study.

On the next finer scale, we observe a differently structured mainland in each of the plots. For the raw DTFE sampling, we see four clearly delineated tongues, the fourth one with lowest mean age smaller than the first three. For the image DTFE sampling, we still see three tongues, although smaller now, and the fourth has disappeared.

We explain these differences by the effective smoothing that differentiates the image DTFE sampling strategy from that of the raw DTFE sampling method. Indeed, the latter picks up the structural details of all four morphological components: clusters, filaments, walls, and cells. The smoothing of the image DTFE sampling is apparently strong enough to remove any traces of the fourth tongue whose high mean density identifies it as the one that used to belong to the cluster morphology.

2.9.2 Density field estimator: DTFE vs. SPH

The second comparison is between a computational procedure defined on the basis of the SPH density estimates, yielding a 3D image representation of the density field on a regular grid, and that based on DTFE density estimates, specifically the image DTFE method. SPH or kernel density estimate has been the standard method of density estimate across various disciplines(Monaghan 2005; Fasy et al. 2013; Phillips et al. 2013; Bobrowski et al. 2014).

The righthand frame of figure 2.21 depicts the dimension 0 persistence intensity diagram based on the SPH density estimate. We observe the same trends as seen in the central diagram, that corresponding to the image DTFE sampling. It displays the same three small tongues, with a trace of a fourth tongue with very high mean density. The latter is likely an artifact of the partially deteriorated clusters. There is not a trace of a fourth tongue seen in the raw DTFE sampling method persistence diagram at low density values, that represent the clusters.

There are some minor differences between the SPH and DTFE persistence diagrams. For instance, SPH density estimates may appear to correspond to walls of a lower persistence than those found in the DTFE density fields.

2.9.3 Summary

Overall, the tests indicate the use of different field sampling strategies or field estimators do not introduce or remove features in the persistence diagrams. Overall, the three frames in figure 2.21 display the same characteristics, although there are also differences. The impact of differences in the density estimator, DTFE or SPH, is considerably smaller than that of the sampling strategy. This appears mainly to be a result of the effective smoothing involved with the field representation on a regular grid in the image DTFE sampling and the SPH method. By contrast, the raw DTFE sampling retains all aspects of the density field traced by the discrete point sample.

2.10 Conclusions and Discussion

This chapter introduces the use of tools emanating from Morse theory, homology and persistence homology for the topological analysis of the mass distribution in the Cosmos. The focus is on the analysis of heuristic models that mimic certain aspects of the cosmic mass distribution. This is motivated from the angle to understand and expose the connection between the salient features of the models vis-a-vis their manifestation in the topological characteristics.

In Section 2.4, we analyze the topology of random, featureless distributions, taking up the particular case of poisson distributed particles, through persistence and Betti numbers. In the process, we characterize the topology of shot-noise as a benchmark, paving the way for a method to estimate the contribution of signals arising due to the presence of shot noise in generic distributions with real features (Section 2.8).

In Section 2.5, we analyze the topology of the pure Voronoi element models. These models are characterized by the presence of exclusively cluster-like, filament-like or wall-like distributions, resembling the different morphologies as seen in the cosmic web (Bond et al. 1996). We show that dominating presence of different morphologies is reflected in the features of the *intensity maps* (averaged persistence diagrams) in different dimensions. Specifically, clustered distributions are characterized by high persistent isolated clouds in 0-dimensional diagrams, loopy distributions characterizing highly filamentary distributions are reflected in a similar high persistence cloud in the 1-dimensional intensity maps, while predominantly wall-like distributions denoting the presence of well formed voids are characterized by isolated high persistence clouds in the 2-dimensional intensity maps.

Section 2.6 analyzes the topology of the multi-scale fractal Soneira-Peebles model (Soneira & Peebles 1978). We demonstrate therein that the levels of hierarchy are well manifested in the intensity maps as well as the graphs of Betti numbers.

In Section 2.7, we analyze the topology of the heuristic Voronoi evolution models. The models seek to mimic the evolution of mass distribution in the Universe, gradually progressing from a stage where most of the matter is confined to the field, to a stage where most of the matter is confined to clusters and filaments. We demonstrate that the topological characteristics of the different morphologies present in the distribution – clusters, filaments and walls – are well segregated in the intensity maps and the graphs of Betti numbers.

In Section 2.9, we analyze the topological characteristics of a single Voronoi evolution model, where the density is computed through different prescriptions, specifically the density computed using DTFE (van de Weygaert & Schaap 2009b) on the raw particle distribution, density computed on a regular grid using DTFE, and density computed on a regular grid using SPH (Hernquist & Katz 1989; Monaghan 2005). We demonstrate that the significant topological properties of the model are retrieved irrespective of the density estimator in use.

We conclude, based on our preliminary investigations on heuristic models, that the intensity maps and Betti numbers are excellent indicators of the presence of multiple morphologies as well as the presence of hierarchies in the distribution. In addition, we can design methods based on these prescriptions to segregate topological noise from real signals. The topological methods are also fairly robust in extracting

the real signals irrespective of the choice of density estimators. We, therefore propose their use as a new standard in the characterization of the cosmic mass distribution.

3

Persistence and Homology of Gaussian random fields: On the hierarchical topology of the primordial field

— Pratyush Pranav—

Every man is more than just himself; he also represents the unique, the very special and always significant and remarkable point at which the world's phenomena intersect, only once in this way, and never again. That is why every man's story is important, eternal, sacred; that is why every man, as long as he lives and fulfills the will of nature, is wondrous, and worthy of consideration. In each individual the spirit has become flesh, in each man the creation suffers, within each one a redeemer is nailed to the cross.

– *Hermann Hesse, Demian: Die Geschichte von Emil Sinclairs Jugend*

3.1 Introduction

A Gaussian random field is an example of a spatial stochastic process. Given a spatial location *s*, a *Gaussian random field* is a random function X(s) on R^3 such that, when restricted to any finite set, one has a multivariate normal distribution.

Gaussian random fields play a key role in describing and modeling the fields arising in cosmology. As a telling example, in the simplest form, the inflationary theories predict the primordial perturbations to be a Gaussian random field (Guth & Pi 1982). The primordial perturbations are a result of the quantum fluctuations, plated during the early inflationary era, in an otherwise homogeneous and isotropic Universe (Peebles 1980; Guth & Pi 1982). The temperature fluctuations in the observable cosmic microwave background, which follow the pattern of the quantum perturbations from the inflationary era, also suggest that the character of the primordial perturbations is that of a homogeneous and isotropic Gaussian random field to high accuracy (Smoot et al. 1992; Bennett et al. 2003; Spergel et al. 2007; Komatsu et al. 2010; Planck Collaboration et al. 2015).

Skipping the description at sub-linear scales, where the details of galaxy formation process become important, the density in the large-scale Universe reflects the characteristics of the primordial fluctuation field. This is because the structures in the Universe emerge from the primordial fluctuation field, the imprints of which are preserved at very large scales. Indeed, if one smooths the cosmic fields over scales much larger than the correlation length, one expects to recover the primordial fluctuation field. As a result, we also expect the Gaussian random fields to model the cosmic fields in the regimes at the later epochs (Adler 1981; Bardeen et al. 1986; Gott et al. 1986, 1989, 2009; Park et al. 2013).

Due to the central role it plays in describing a multitude of fields of interest that arise in cosmology, the characterization of Gaussian random fields has been an important focal point in cosmological studies (Doroshkevich 1970; Bardeen et al. 1986; Gott et al. 1986; Bertschinger 1987; Scaramella & Vittorio 1991; Schmalzing & Buchert 1997; Matsubara 2010). A Gaussian random field is fully specified by its power spectrum. As a result, the determination and characterization of the power spectrum of the theoretical models as well as observational data has been one of the main focal points in the analysis of the primordial fluctuation field as well as the large scale cosmic fields. Within this, the analysis of the power spectrum of the Cosmic Microwave background has played the key role in the study of the characteristics of the primordial fluctuation field (Scaramella & Vittorio 1991; Eisenstein & Hu 1999; Oh et al. 1999; Efstathiou & Bond 1999; Seljak & Zaldarriaga 1999b; Peterson et al. 1999; Knox & Page 2000; Wilson et al. 2000; Grainge et al. 2003; Durrer et al. 2003; Tristram et al. 2005; Hazra et al. 2013).

Topology is the branch of mathematics that is concerned with the properties of space that are preserved under continuous deformations including stretching (compression) and bending, but not tearing or gluing. It also includes invariance of properties such as connectedness and boundary. Topological studies of the cosmic density fields have also been very insightful in understanding their properties. This is because a topological description provides an insight into the the prominence of topological features, and the global structure as well as the connectivity character-

istics of the field. The analysis of the connectivity characteristics help to develop a better understanding of how cosmic structures emerge, and subsequently interact and merge with neighboring features, as structure formation evolves in the Universe. Because topology is insensitive to continuous deformations like stretching and bending, topological measures are also expected to be relatively insensitive to systematic effects such as non-linear gravitational evolution, galaxy biasing, and redshift-space distortion (Park & Kim 2010). As a result, one expects the topology of the primordial field and the large-scale linear field at the later epochs to be similar (Gott et al. 1986, 1989, 2009; Park et al. 2013). Formally, this is only true as long as there is no shell-crossing in the evolving mass distribution. When shell-crossing occurs the field may change its connectivity characteristics, and as such there may be a net difference between the topological properties of the field before and after shell-crossing.

The topological characterization of the models of cosmic mass distribution has also been a focal point of many studies (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986; Canavezes et al. 1998; Canavezes & Efstathiou 2004; Pogosyan et al. 2009; Choi et al. 2010; Park & Kim 2010). The early studies of the topological characteristics of the cosmic mass distribution were based on the evaluation and analysis of the genus and the Euler characteristic of the iso-density surfaces for models as well as the observational data (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986; Canavezes et al. 1998; Canavezes & Efstathiou 2004; Pogosyan et al. 2009; Choi et al. 2010; Park & Kim 2010; Park et al. 2013). Gott and collaborators (Gott et al. 1986; Hamilton et al. 1986) introduced the use of the genus as a function of density threshold to characterize the cosmic mass distribution. Refer to Section 1.5.1 and Appendix A for a detailed definition and illustrative examples of the genus and the Euler characteristic.

Genus or Euler characteristic studies have played a key role in the topological studies of the cosmic density fields. Of fundamental importance in this respect has been the realization that the expected value of the genus in the case of Gaussian random fields for iso-density surfaces as a function of density threshold has an analytic closed form expression (Adler 1981; Bardeen et al. 1986; Gott et al. 1986):

$$g(\nu) \;=\; -\frac{1}{8\pi^2}\left(\frac{\langle k^2\rangle}{3}\right)^{3/2}(1-\nu^2)e^{-\nu^2/2}. \tag{3.1}$$

In this expression, $\nu = \delta/\sigma$ is the dimensionless density threshold, specifying the number of standard deviations that the density contrast at a particular location differs from the mean. Important to note that functional form of the genus is independent of the specification of the power spectrum for Gaussian fields, and is a function only of the dimensionless density threshold ν. The contribution from power spectrum is restricted to the amplitude of the genus curve through the quantity $\langle k^2\rangle$, which is related to the second moment of the power spectrum.

Later, more discriminative information became available with the introduction of Minkowski functionals (Mecke et al. 1994; Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). There are $(d+1)$ Minkowski functionals, $Q_k((k = 0,\ldots,d)$, defined for a d-dimensional manifold (Mecke et al. 1994; Schmalzing &

Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). Predominantly geometric in nature, the first four Minkowski functionals ($d \leq 3$) are respectively the volume, surface area, integrated mean curvature or total contour length, and the Gaussian curvature. Refer to Section 1.5.2 for a detailed exposition on the Minkowski functionals.

Statistics based on the Minkowski functionals have played a key role in the the topological and morphological studies of the cosmic density fields. For Gaussian random fields, the expected value of the first four Minkowski functionals of the excursion sets have known analytical expressions (Tomita 1993; Schmalzing & Buchert 1997). Amongst others, this makes them an ideal tool for validating the hypothesis of initial Gaussian conditions through a comparison with the observational data.

The genus, the Euler characteristic and the of the Minkowski functionals, have been extremely instructive in gaining an understanding about the connectivity as well as the structural patterns pervading in the Universe. However, the topological information represented by the genus and the Euler characteristic is limited. A full description should involve a considerably more extensive palette of topological characteristics, than what the Minkowski functionals may supply. The foremost compelling reason to look beyond the available topological descriptors is the observation that structures in the Universe form and evolve in a hierarchical fashion. Smaller high density structures coalesce together hierarchically to build up larger structures of lower density. A topological description through the Minkowski functionals is not equipped to address this aspect of the cosmic mass distribution. A formalism capable of expressing topology in a hierarchical fashion would present an interesting and powerful extension to the existing topological descriptors.

In view of the above observations, this chapter seeks to present a topological analysis of Gaussian random fields through *homology* (Munkres 1984; Edelsbrunner & Harer 2010; Adler & Taylor 2010) and *persistence* (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Carlsson et al. 2005; Carlsson & Zomorodian 2009; Carlsson 2009; Edelsbrunner & Harer 2010). For a detailed exposition on homology and persistence, refer to Appendix A, as well as Chapter 1 (Section 1.6.1 and Section 1.6.3. The topological properties of probability distributions on a manifold, and and the topological properties of random complexes [1] have been an active area of research recently (Adler et al. 2010; Bobrowski & Strom Borman 2010; Feldbrugge et al. 2015; Adler et al. 2014; Bobrowski & Kahle 2014; Bobrowski & Mukherjee 2015). Homology and persistence have played a key role in these studies. For a survey on the topology of random geometric complexes, see Bobrowski & Kahle (2014).

The formalism of persistence is inherently hierarchical in nature. This may be inferred from examining the topological changes that occur in 0-dimension. One may recall that the 0-dimensional holes are isolated objects (islands). The death of a 0-dimensional hole is equivalent to the merger of two isolated objects. In this sense, the information about persistence of the 0-dimensional holes represents the information about the distribution of the peaks of the density field, as well as the merger processes that occur to form larger and larger objects, as the density threshold de-

[1] A *random complex* is a simplicial complex constructed from a given random point distribution The point distribution is sampled from a well behaved probability distribution function.

creases. This is not unlike the the build up of structures in the hierarchical structure formation scenarios, where smaller high density objects coalesce together to form larger and larger structures of lower density.

This chapter presents a numerical investigation of the topological properties of Gaussian random fields through persistence and homology. In a related paper (Park et al. 2013), we present a preliminary investigation on the homology of Gaussian random fields, as described by the Betti numbers. A semi-analytic theoretical framework towards calculating the Betti numbers and persistence of Gaussian fields has been developed recently by Feldbrugge et al. (2015). In this chapter, we extend the study to an elaborate and systematic numerical analysis of persistence and homology of Gaussian random fields.

We introduce the *intensity maps* as ensemble averages of the persistence diagrams of stochastic processes. Keeping model comparison as an important objective in mind, we also introduce the *difference maps* and the *ratio maps*. First, we delve into the persistence topology of 1D Gaussian functions. This is done in order to relate the visual features of the density distribution to the topological description, and assess how the topological description reflects the features of the density distribution. We establish that the features of the persistence diagram are a direct reflection of the singularity structure of the function. Having assessed the persistence topology of the 1D functions, we delve into an analysis of the 3D fields. We do this for the power law models and the LCDM model. We demonstrate that the features in the intensity and the difference maps depend on the choice of the model. The indication is that the the intensity and the difference maps are highly sensitive to the parameters of the model, and may therefore be used to discriminate between various models.

We quantify the intensity maps by decomposing them into marginal and cumulative distributions as a function of the mean-density and persistence of the topological holes. In this context, we establish that the distribution functions show a characteristic dependence on the index of the power spectrum. We also show that the marginal distribution of the mean density of the holes follow a near Gaussian distribution. The marginal distribution as a function of the persistence of the holes indicates a Poisson distribution.

Subsequently, we delve into an investigation of the Betti numbers of Gaussian random fields. We compare the Betti number curves with the Euler characteristic and show that while the Euler characteristic is insensitive to the index of the power spectrum, the Betti numbers show a systematic dependence on it. We also present a brief analysis on the Minkowski functionals of the model towards the end. Together, the Minkowski functionals, homology and persistence establish a more comprehensive and detailed picture of the topology and morphology of the models.

We begin by providing a description of Gaussian random fields in Section 3.2. Thereafter, we present a brief description of the topological background in Section 3.3. Section 3.4 through Section 3.7 present the main results of this paper. We conclude the chapter by discussing the results in Section 3.8.

3.2 Gaussian random field

A *random variable* is a variable whose value is subject to probabilistic variations. Random variables model the outcome of *stochastic processes*. The values of a random variables are not fixed, but rather drawn from a *distribution function*. The most common is the *Gaussian distribution function*, given by (Adler 1981)

$$f(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{3.2}$$

where σ is the standard deviation of the variable. We implicitly assume that the mean or expectation of the variable is zero.

A generalization of a random variable is a *random field*, which is a collection of random variables. A random field can be specified by the m-point joint distribution function

$$P[f_1,\ldots,f_m]df_1\ldots df_m. \tag{3.3}$$

A random field is a *Gaussian random field* if the set of constituent random variables are all drawn from Gaussian distributions. The m-point joint probability distribution function for a Gaussian random field is a multi-variate Gaussian, given by (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986)

$$P\left[f_1,\ldots,f_m\right] df_1\ldots df_m \quad = \quad \frac{1}{(2\pi)^N (detM)^{1/2}} \cdot \exp\left(-\frac{\sum \Delta f_i (M^{-1})_{ij} \Delta f_j}{2}\right) df_1\ldots df_m, \tag{3.4}$$

where, M^{-1} is the inverse of the $m \times m$ covariance matrix M_{ij}.
The equation is in the normalized form, such that the integral of $P\left[f_1,\ldots,f_m\right] df_1\ldots df_m$, over all $f \in \mathbb{R}^N$, is equal to 1. In the above expression,

$$\begin{aligned} \Delta f_i &= f_i - \langle f_i \rangle \\ M_{ij} &= \langle \Delta f_i \Delta f_j \rangle \end{aligned} \tag{3.5}$$

The matrix M^{-1} is the inverse of the $m \times m$ covariance matrix M_{ij}. The angle bracket denotes the ensemble average of the product. In effect, M is the generalization of the variance of the 1-point normal distribution, and $M = [\sigma_0^2]$ for the case $m = 1$.

3.2.1 Properties of Gaussian random fields: correlation function and power spectrum

Equation (3.4) shows that a Gaussian random is fully specified by the second order moment, via the autocorrelation function $\xi(\mathbf{r})$, encoded through the covariance matrix M. The latter expresses the correlation between the density values at any two points $\mathbf{r_1}$ and $\mathbf{r_2}$ at a distance $\mathbf{r} = \mathbf{r_1} - \mathbf{r_2}$.

If the field is homogeneous and isotropic, the correlation only depends on the

absolute distance between the points $r = |\mathbf{r}|$, such that

$$\xi(r) \;=\; \xi(|\mathbf{r}|) \;\equiv\; \langle f(\mathbf{x}) f(\mathbf{x}+\mathbf{r}) \rangle \,. \tag{3.6}$$

In other words, the entries in the matrix are the values of the *autocorrelation function* for the distance between the points: $M_{ij} = \xi(r_{ij})$, with

$$r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| \,. \tag{3.7}$$

To appreciate the contribution from different scales, the structure of a Gaussian field is more transparently characterized by the power spectrum $P(k)$. It is the Fourier transform of the autocorrelation function, given by

$$\langle \hat{f}(\mathbf{k}) \hat{f}(\mathbf{k}') \rangle \;=\; (2\pi)^{3/2} \, P(k) \, \delta_D(\mathbf{k} - \mathbf{k}') \,, \tag{3.8}$$

where, $\delta_D(\mathbf{k})$ is the Dirac delta function. This implies that the knowledge of the power spectrum alone is sufficient to fully characterize a Gaussian random field. The contribution of power at a particular scale to the total variance of the density field is specified by the relation

$$\sigma^2 = \int_0^\infty dk \, k^2 P(k) = \int_0^\infty d\ln k \, k^3 P(k). \tag{3.9}$$

such that, $k^3 \, P(k)$ is the contribution of the power spectrum per unit logarithmic bin to the total variance of the density field.

This chapter investigates the topological properties of zero-mean Gaussian random fields. They are specified as density values on a regularly spaced cubic grid. We focus on the models with cosmologically relevant power spectra – power-law power spectra and the LCDM power spectrum. In all, we investigate six models – the LCDM model, and five power-law models with different spectral indices.

Power-law power spectrum. The power-law power spectrum is a generic class of spectrum, specified by the spectral index n

$$P(k) = A_n \, k^n. \tag{3.10}$$

The case when $n = 1$, the Harrison-Zel'dovich spectrum, is the predicted spectrum of the primordial density perturbations (Harrison 1970; Peebles & Yu 1970; Zeldovich 1972). This is because of its property of scale invariance, which makes it natural choice for the primordial power spectrum (Dunkley et al. 2009; Komatsu et al. 2011; Planck Collaboration et al. 2013b). The measured spectrum of the primordial perturbations is very close to it, $n \sim 0.96$ (Dunkley et al. 2009; Komatsu et al. 2011; Planck Collaboration et al. 2013b). It is worthwhile noting here that certain inflationary theories also predict $n \sim 1$ for the power spectrum. The top-left and the bottom-left panels of Figure 1.2 plots the power spectrum $P(k)$ and $k^3 \, P(k)$ for the power law models.

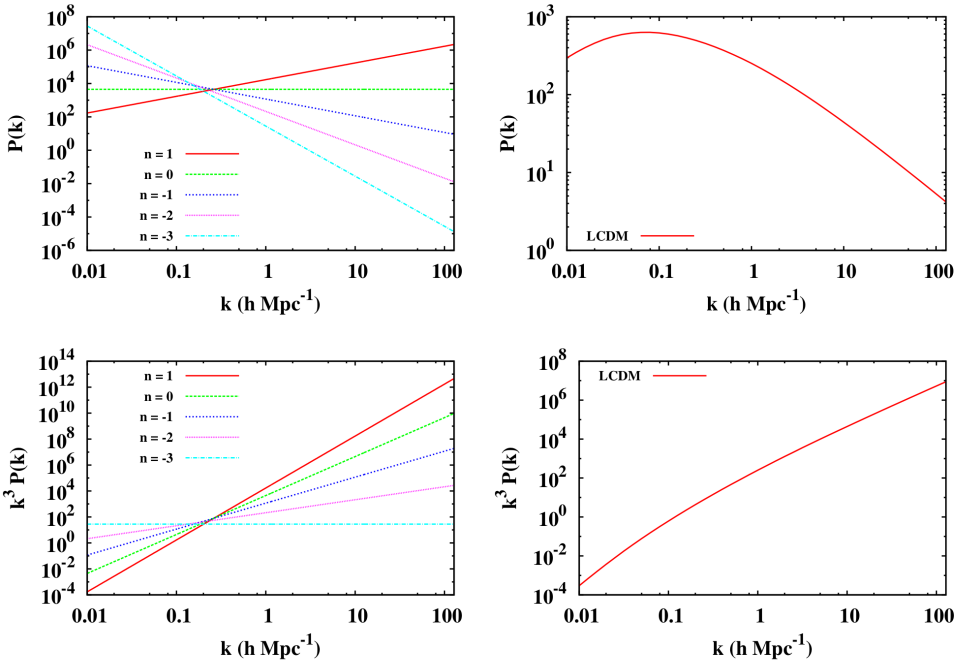


Figure 3.1 The scaled power spectrum $P(k)$, as well as power spectrum per unit logarithmic bin $k^3\,P(k)$. Graphs are presented for the different spectral indices of the power-law model, as well as the LCDM model. The spectra are scaled such that different models have the same variance of the density fluctuations, when filtered with a top-hat filter of radius $8h^{-1}$Mpc.

LCDM power spectrum. The LCDM power spectrum stems from the standard concordance model of cosmology. It fits the measured power spectrum of the cosmic microwave background as well as the power spectrum measured in the nearby large scale Universe to high accuracy. It is given by (Eisenstein & Hu 1999; Hu & Eisenstein 1999)

$$P(k) \propto T^2(k) P_0(k), \tag{3.11}$$

where $P_0(k)$ is the primordial power spectrum and $T(k)$ is the transfer function. The transfer function is an expression of physical processes acting on the fluctuations as they enter the horizon. Its shape can be inferred by evaluating the evolving processes, through the Boltzmann equation (Seljak & Zaldarriaga 1999a). A good numerical fit is given by (Eisenstein & Hu 1999; Hu & Eisenstein 1999)

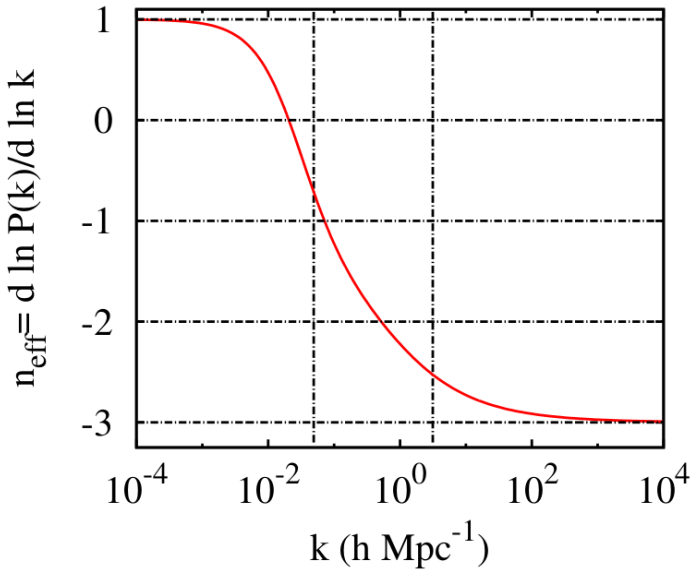


Figure 3.2 The effective spectral index for the LCDM power spectrum within the simulation box analyzed in this chapter. The effective spectral index of the CDM power spectrum is a function of the wave number k. The vertical dashed lines correspond to the wavenumbers associated with the fundamental and the Nyquist mode of the box. The horizontal dashed lines are drawn for comparison with the index of power spectrum for the power-law models.

$$P_{CDM}(k) \propto \frac{k^n}{[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4]^{1/2}} \times \frac{[ln(1+2.34q)]^2}{(2.34q)^2}, \tag{3.12}$$

$$q = k/\Gamma,$$

$$\Gamma = \Omega_m h \exp\left\{-\Omega_b - \frac{\Omega_b}{\Omega_m}\right\}.$$

In the above expression, Ω_m and Ω_b are the total matter density and baryonic matter density respectively. Γ is referred to as the shape parameter. Inserting the most recent observed values of Ω_b and Ω_m, $\Gamma \sim 0.21$. The top-right and the bottom-right panels of Figure 3.1 plot $P(k)$ and $k^3\,P(k)$ for the CDM power spectrum. It can be seen that the spectral density per logarithmic bin increases with increasing wavenumbers or decreasing wavelengths. As a result, the LCDM power spectrum shows very prominent small scale features.

Locally, the spectrum resembles a power-law, with the spectral index $n_{eff}(k)$ showing a dependence on the scale k, through the relation

$$n_{eff}(k) = \frac{\mathrm{d}\ln P(k)}{\mathrm{d}\ \ln k}. \tag{3.13}$$

Figure 3.2 plots the effective spectral index $n_{eff}(k)$ for the CDM power spectrum as a function of scale k. In the asymptotic limit of small and large k, the limits of $n_{eff}(k)$ are well defined. At very large scales, its behavior tends towards a power-law with index $n = 1$, as can be seen in the plot. At small scales, the LCDM power spectrum behaves like a power-law power spectrum with index $n = -3$. The vertical dashed lines in the plot correspond to the fundamental and the Nyquist mode of the simulation box respectively. It is evident from the plot that the effective index of the model varies steeply between $n_{eff} \sim -0.5$ to $n_{eff} \sim -2.5$. At the lower limit, the Nyquist mode of the box corresponds to the scales of milkyway like galaxies. At the other end the fundamental mode of the box corresponds to wavelengths well beyond the scales at which the Universe appears homogeneous.

3.2.2 Normalization of the power spectrum

The amplitude of the power spectrum is not predicted by the fundamental physical theories. It has to be determined from observations of the amplitude of mass fluctuations in the Universe. One of the means to calculate its value is by equating the theoretically calculated variance of the density field in the Universe to the observationally determined value, filtered by a top-hat filter of radius $8h^{-1}$Mpc

$$\sigma_8(obs) = A \int \frac{d\vec{k}}{(2\pi)^3} P(k) \hat{W}^2_{TH}(k). \tag{3.14}$$

where, $\hat{W}_{TH}$ is the top-hat filter given by

$$\hat{W}_{TH}(x) = \frac{3}{x^3}(sinx - xcosx). \tag{3.15}$$

The left-hand side of Equation 3.14 is the rms of the observed value of density fluctuations. The integral on the right hand side of the equation is solved numerically. The ratio of the observed σ_8 to the calculated integral gives the value of the amplitude A of the power spectrum. For the models in this chapter, Figure 3.1 presents the graph of the normalized power spectra for the various power-law models as well as the CDM model.

It is important to note that the power-law power spectrum has no characteristic scale, and diverges asymptotically for either small or large values of k, depending on the choice of the power spectrum. As a result the integral in Equation 3.14 also diverges for the power-law models. In order to remedy this, the usual practice is to set the limits of integration to the fundamental mode and the Nyquist frequency of the of the simulation box. The fundamental mode k_{fund} and the Nyquist mode k_{Nyq}

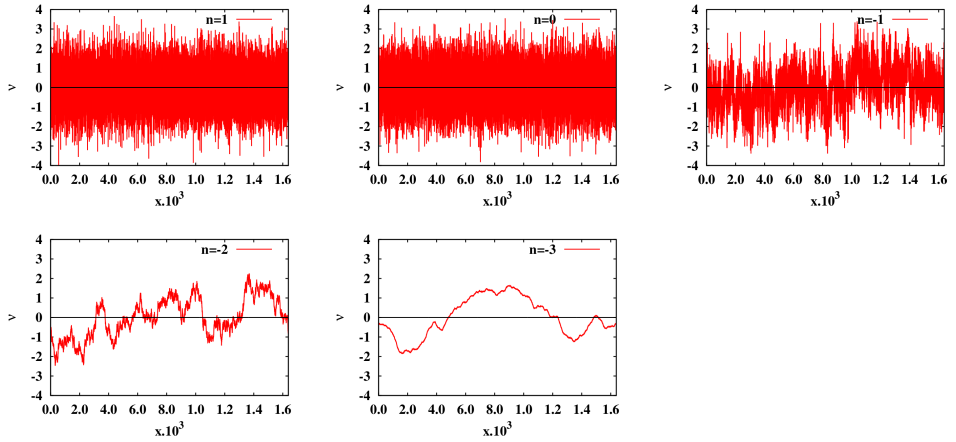


Figure 3.3 Density fluctuations in a random realization of a 1D Gaussian random field with a power law power spectra. The spectral index are $n = 1, 0, -1, -2$ and -3. For the positive spectral index there is more power at small scales. As a result, the appearance of the field is spiky. As the spectral index decreases, the power shifts towards larger scales. As a result, there are more prominent large scale features gradually. The small scale features appear as tiny wiggles, due to decreasing power at those scales.

are inversely proportional respectively to the box size L, and twice the grid size, l. For the models discussed in this chapter, $k_{fund} = 2\pi/128h^{-1}\mathrm{Mpc} \sim 0.049hMpc^{-1}$ and $k_{Nyq} = 2\pi/2h^{-1}\mathrm{Mpc} \sim 3.14hMpc^{-1}$.

3.2.3 Model realization

We examine 1D and 3D models of Gaussian random fields. In both the cases, the models are constructed in a simulation box of side $128h^{-1}$Mpc with a grid resolution of $1h^{-1}$Mpc. For the 1D case, we analyze models specified by a power law power spectrum. The value of the spectral indices are $n = 1, 0, 1, -2$ and -3. For the 3D case, we analyze models specified by a power law power spectrum, with indices $n = 1, 0, 1, -2$ and -3. In addition, we also analyze a model specified by the LCDM power spectrum. All the models are smoothed with a Gaussian kernel of scale $R_f = 2h^{-1}$Mpc. The results presented for all the models are averaged over 100 realizations. The density fluctuation δ at each location is normalized by the square root of the variance of the density fluctuation σ_0, such that the dimensionless density threshold ν is given by

$$\nu = \delta/\sigma_0. \tag{3.16}$$

The results in this chapter are quoted in terms of the dimensionless threshold ν.

Figure 3.3 presents the density fluctuations in a random realization of a 1D Gaussian random field. Figure 3.4 shows 2D slices of a single realization of 3D Gaussian

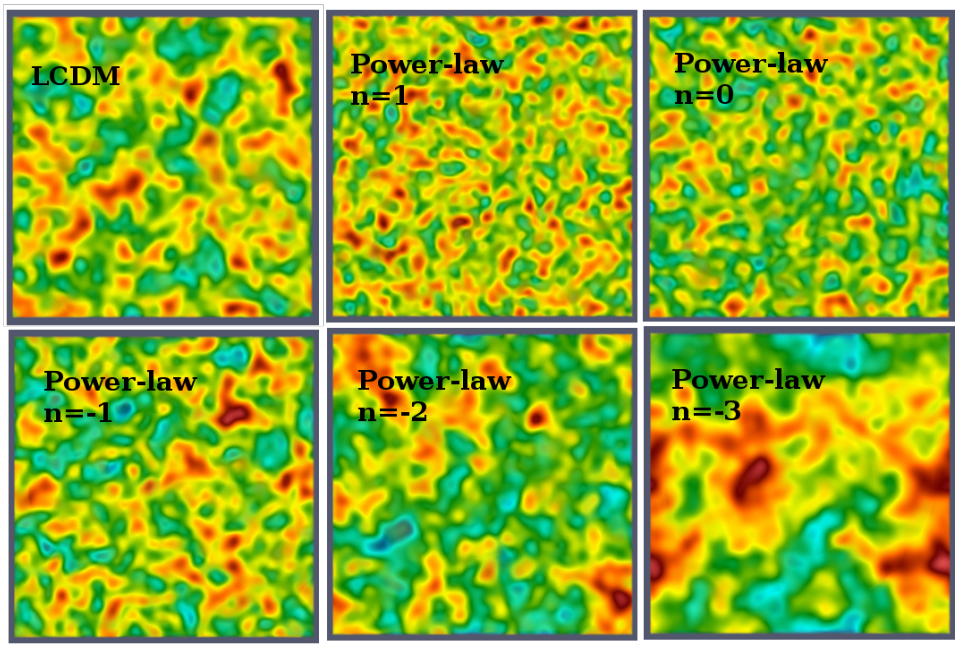


Figure 3.4 2D slices of a single realization of 3D Gaussian random field models investigated in this thesis. The models are constructed in a simulation box of side $128h^{-1}$Mpc with a grid resolution of $1h^{-1}$Mpc. Subsequently, it is smoothed with a Gaussian kernel of scale $R_f = 2h^{-1}$Mpc. The top left panel shows a realization of the LCDM power spectrum. The rest of the panels show realizations of power-law power spectra with spectral indices $n = 1, 0, -1, -2$ and -3. As we go from positive to progressively negative spectral indices, the structures become visibly larger. This indicates increasing power at larger scales with decreasing spectral index.

random field. The top left panel shows a realization of the LCDM power spectrum. The rest of the panels show realizations of power-law power spectra. For these models, there is relatively more power at the small scales for a higher spectral index, in comparison to a lower spectral index. This can be inferred from the shape of the curves for $k^3\,P(k)$ in the lower left panel of Figure 3.1. As a result, the appearance of the field is spiky. As the spectral index decreases, the power shifts towards larger scales. As a result, there are more prominent large scale features gradually. The small scale features appear as tiny wiggles, due to decreasing power at those scales.

3.3 Topology

In this section, we give an informal presentation on the theory of homology and persistence, and the concepts essential to its formulation. As Gaussian random fields are an example of a stochastic process, we briefly revisit intensity maps as an empirical description of persistence. Keeping our mind focused on model discrimination, we also introduce the difference and ratio maps as new concepts. Then we discuss

the Betti numbers and argue that they are emergent from the description of persistence. Subsequently, we define and describe the genus and the Euler characteristic. Thereafter, we establish that persistence diagrams, Betti numbers and genus or Euler characteristic strictly contain decreasing amount of information. We refer the reader to Appendix B for technical details on the computational procedure involved.

3.3.1 Homology

Homology is a mathematical formalism for specifying in a quantitative and unambiguous manner about how a manifold is connected, through assessing its boundaries and cycles (for a detailed and standard discourse on the subject, see Munkres (1984); Edelsbrunner & Harer (2010); also see Appendix A). Homology groups provide a mathematical language for describing the holes in a topological space. A d-manifold can be composed of topological holes of 0 up to $(d-1)$ dimensions. The holes in the first three dimensions have intuitive interpretations. A 0-dimensional hole is a *gap* between two isolated independent objects. A 1-dimensional hole is a *tunnel* through which one can pass in any one direction without encountering a boundary. A 2-dimensional hole is a *cavity* or *void* fully enclosed by a 2-dimensional surface. Alternatively, the holes can be defined by the *cycles* that form their boundary. Two 0-cycles form the boundary of the the gap (0-dimensional hole) between them. A 1-cycle bounds a tunnel, and a 2-cycle bounds a void. The collection of all p-dimensional cycles is the p-th homology group $\mathbb{H}_p$. The rank of this group is denoted by the *Betti numbers* β_p, where $p = 0, \ldots, d$ (Betti 1871; Edelsbrunner & Harer 2010). β_0 counts the number of independent components, β_1 counts the number of loops enclosing the independent tunnels and β_2 counts the number of shells enclosing the independent voids. Also refer to Appendix A for more details on homology.

3.3.2 Topology and singularities: Morse functions

Morse theory (also see Appendix A.4) studies the change of topology of the manifold $\mathbb{M}$ as induced by the superlevel-sets of the function f. The *superlevel sets* of a field are defined as the regions

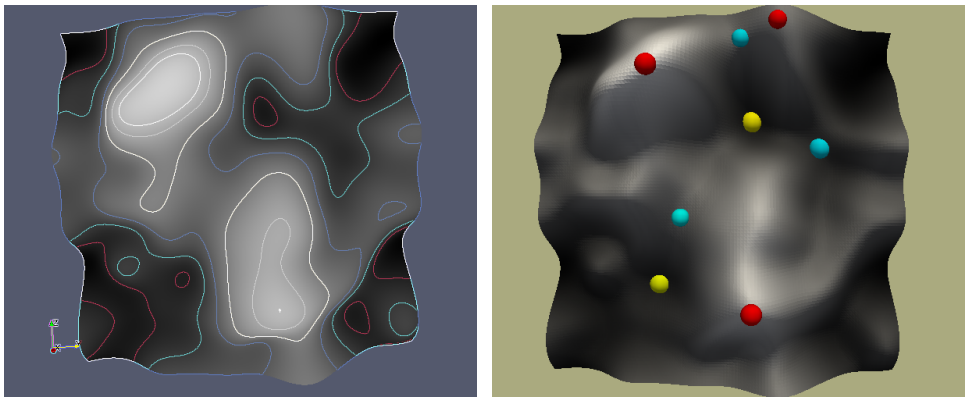
$$\begin{aligned} \mathbb{M}_\nu &= \{\mathbf{x} \in \mathbb{M} \mid f_s(\mathbf{x}) \in (f_\nu, \infty]\} \\ &= f_s^{-1}(-\infty, f_\nu]. \end{aligned} \tag{3.17}$$

In other words, they are the regions where the smoothed density is greater than or equal to the threshold value $f_\nu = \nu\sigma_0$, with σ_0 the dispersion of the density field.

The key property is that the topology of the manifold changes only when passing through a critical point of f. A critical point is the point where the gradient of the function vanishes, i.e.

$$\nabla f = 0. \tag{3.18}$$

The type of the critical point is decided by the number of negative indices in the eigenvalues of the Hessian of the function at that location, which is the matrix of the



(a) (b)

Figure 3.5 (a) Face on view of a 2D random field. A range of level sets are presented as contours of different colors. (b) Surface view of the same. Maxima are marked with red balls, saddles with cyan balls, and minima with yellow balls.

partial double derivatives of the function, given by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}.$$

A 1D function can only have two kinds of singularities: a maximum or a minimum. A negative eigenvalue for the Hessian denotes a local maximum, and a positive eigenvalue denotes a local minimum. A 2D function can have three kinds of singularities: maxima, minima and saddles. In the case of a maximum, the eigenvalue of the Hessian has two negative quadratic terms. For a saddle point, there is a positive and a negative term. For a minimum, there are two positive terms in the eigenvalue. In 3D, there are four kinds of critical points: maxima, minima and two different kinds of saddle points. A maximum has three negative eigen values, and the minimum has three positive eigen values. The different kind of saddles have two positive (or negative) and one negative (or positive) eigen values. The type of the saddle is determined by the particular combination (see Section 1.6.2 for the definition and illustration of the critical points in 2D and 3D).

3.3.3 Hierarchical topology: Persistence

In Section 3.3.2, we noted the relation between the critical points of a function with the topological changes it induces in a manifold. In this section, we use the notions described above to sketch an intuitive understanding of persistence homology (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Carlsson et al. 2005; Carlsson & Zomorodian 2009; Carlsson 2009; Edelsbrunner & Harer 2010).

Persistence is a hierarchical extensison of homology. Having deep connections with Morse theory (Milnor 1963), at the heart of the formalism of *persistence* is the key observation is that the topology of the manifold changes only when passing

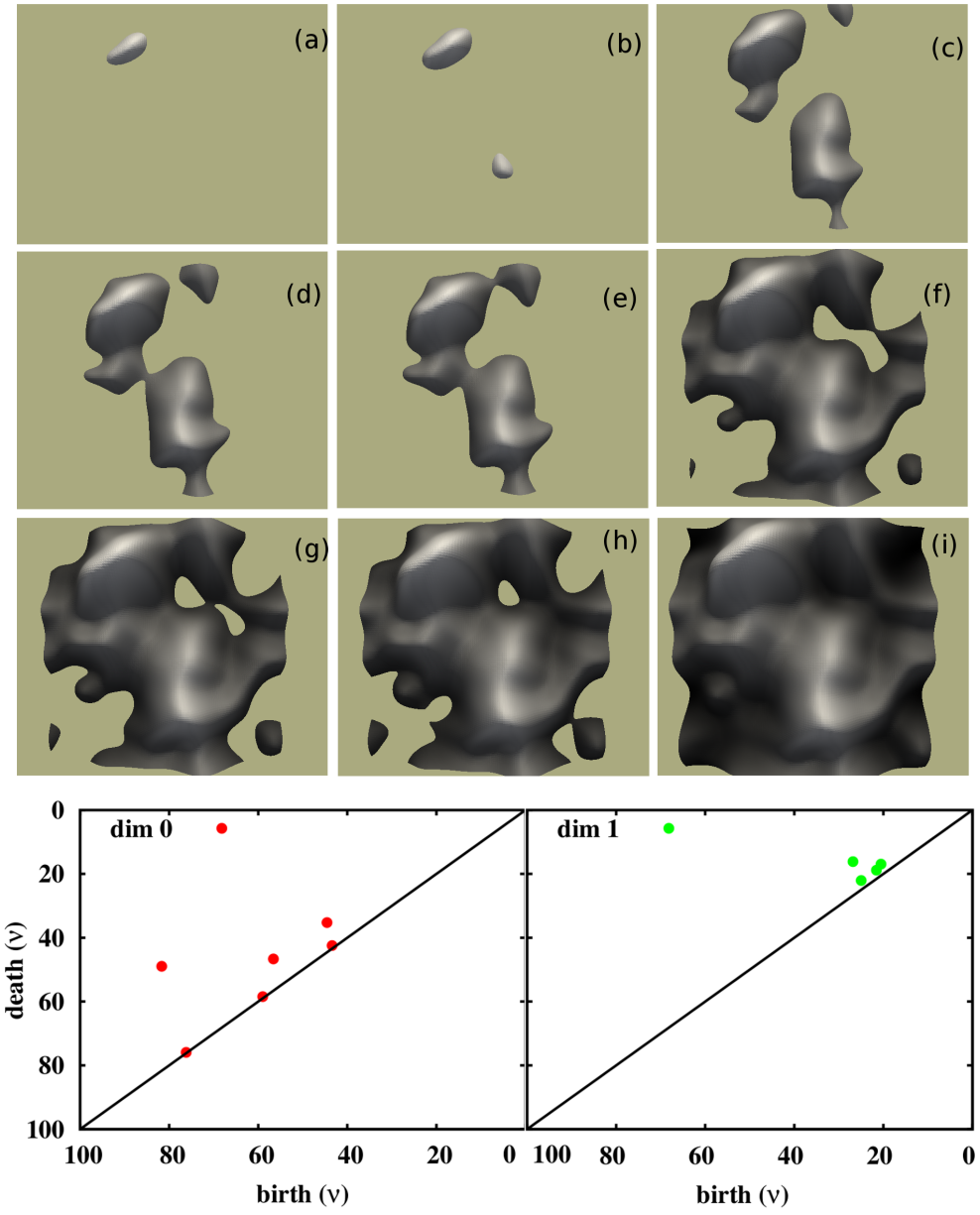


Figure 3.6 Persistence and field singularity structure. The process of birth and death as we grow the superlevel sets by decreasing the density threshold in a given 2-dimensional random field. The events of birth and death in each dimension are recorded separately in the persistence diagrams. Bottom-left: 0-dimensional diagram, bottom-right: 1-dimensional diagram.

through a critical point. More specifically, the addition of a p-critical point can result in either the *birth* of a p-dimensional hole or the *death* of a $(p-1)$-dimensional hole (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Edelsbrunner & Harer 2010). Central to the formulation of persistence is the necessity to track the birth and death events, as one changes the superlevel or sublevel sets. Tracking the topological changes through changing superlevel sets, is equivalent to tracking it through changing sublevel sets, by noting that β_0 of the superlevel sets is equivalent to β_2 of the sublevel set, and vice-versa. Theoretically, there are infinitely many levels of density as one decreases the value from the highest to the lowest. But fortunately, the topology of the manifold only changes while passing a critical point (Edelsbrunner et al. 2002; Edelsbrunner & Harer 2010). This is of key importance because this means that the infinite number of levels of density threshold can be constrained to a finite number, by only having to consider one level in between any two critical points. In addition, each topological hole is associated with two unique function value: $f(c_b)$ associated with the critical point c_b that gives birth to the hole, and $f(c_d)$ associated with the critical point c_d that is responsible for *killing* or filling up the hole. The *lifetime* or *persistence* π (Edelsbrunner et al. 2002; Edelsbrunner & Harer 2010), of the hole is then given by the absolute difference between the death and the birth values associated with the hole

$$\pi = |f(c_b) - f(c_d)|. \tag{3.19}$$

Persistence: an example

Figure 3.5 presents a 2D random field, that we use as an example to illustrate the working of persistence. In panel (a), we present the face on view of the field. A range of level sets are presented as contours of different colors. In panel (b), we present the surface view of the same. Maxima are marked with red balls, saddles with cyan balls, and minima with yellow balls. Figure 3.6, illustrates the working of persistence by tracking the birth and death of islands and tunnels, for growing superlevel sets, for the $2D$ random field show in Figure 1.15. We employ a $2D$ field for ease of visualization. An example in 2D presents a clear intuitive understanding also of the process in 3D.

We trace the growing superlevel set from the top-left panel to the bottom right panel, as a function of decreasing thresholds of f. We only show regions of the manifold that are included in the superlevel set. We trace the change in topology of this region, as the superlevel set grows. In panel (a), we start with a single island, which is also a peak. In panels (b) and (c), we witness the birth of two more islands. These are two peaks which get included in the excursion region. In panel (d), two of the islands merge and we are left with two islands as a result. This merger results in the death of one of the islands. This is the island which was born at a later threshold. This is according to the *elder rule* (Edelsbrunner et al. 2002; Zomorodian & Carlsson 2005; Edelsbrunner & Harer 2010), which dictates that given a choice between killing two components, the one that is born later is killed preferentially. In panel (e), there is another merger of two isolated islands.

In panel (f), the first 1-dimensional hole or a loop is born. It has the appearance

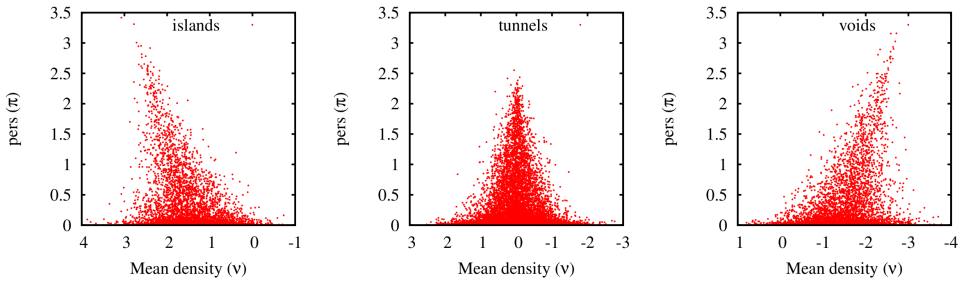


Figure 3.7 Typical persistence diagram for islands, tunnels and voids for a white noise Gaussian random field in 3D.

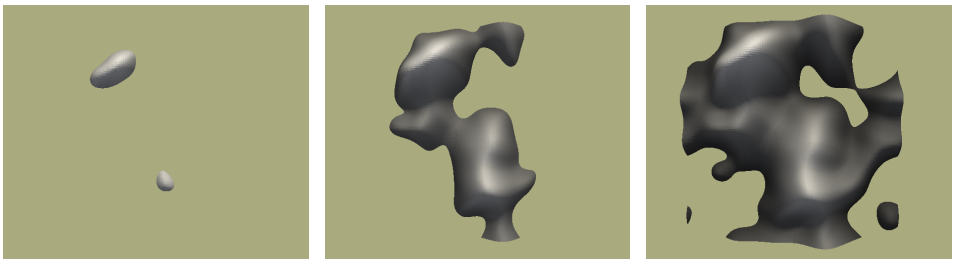
of a lake surrounded by land. In panel (g), this loop splits into two. In panel (h), one of the loops is filled, while the other one still exists. This indicates the death of a 1-dimensional hole. The density value at which a hole is born, and the density value at which it gets destroyed quantify its *life-span*. The life-span is equal to the difference between the absolute values of the density of birth and death. This means that one of the two loops has a higher life span or persistence than the other. It is likely to be a more significant feature than the other. In panel (i), all the holes fill up, and the superlevel set consists of the whole manifold.

3.3.4 Persistence diagrams

Persistence homology is represented in terms of *persistence diagrams* (Edelsbrunner et al. 2002; Edelsbrunner & Harer 2010), which is a collection of dots, each dot associated with a unique topological change in the manifold. There is a diagram for each ambient dimension of the manifold. 0-dimensional diagrams record the merger events of two isolated objects. 1-dimensional diagrams record the formation and destruction of loops, while 2-dimensional diagrams record the birth and death of topological voids. We have introduced a representation of the persistence diagrams that involves a rotation according to

$$b : d \rightarrow \frac{d+b}{2} : d - b. \tag{3.20}$$

In this representation, the horizontal axis $\left(\frac{d+b}{2}\right)$ is the mean-density of the feature. The vertical axis $(d-b)$ is the *persistence*, or *life-span* of the feature. The persistence diagrams corresponding to the birth and death events depicted in Figure 3.6 are presented in in the bottom row. The bottom-left panel presents the persistence diagram for islands, and the bottom-right panel presents it for tunnels. Figure 3.7, presents the typical dot diagrams for islands, tunnels and voids for a random realization of a 3D white noise Gaussian random field.



(a) (b) (c)

Figure 3.8 Figure illustrating the difference between peaks and islands. The left panel illustrates two peaks. They are composed of a single maximum. However, since they are also trivially connected and isolated objects, they double up as islands also. The middle and the right panels illustrate islands with a more complex topology. In the middle panel the island is a connected object, and contains many peaks. In the right panel, the island encloses a loop as well.

3.3.5 Peaks vs. Islands

There is a telling distinction between *peaks* such as described by Bardeen et al. (1986), and the *islands* of our definition. An island is a single connected object. A peak is the location of a local maximum of the function. In general, an island may be marked by many peaks. However, at the higher density thresholds, when no saddle points have yet been introduced in the manifold, there will be necessarily one peak per island. As the threshold is lowered, the number of peaks per island increases. As this happens, the manifold starts developing complex connectivity. This happens because the peaks merge through saddles, and an island may even form tunnels.

As an example, Figure 3.8 illustrates the difference between peaks and islands. The left panel illustrates two peaks. They are composed of a single maximum. However, since they are trivially connected and isolated objects, they can also be classified as islands. The middle and the right panels illustrate islands with a more complex topology. In the middle panel, the island is a connected object, and contains many peaks. In the right panel, the island encloses a loop as well. In this context, we point out that the number of peaks per island, as a function of the density threshold, is a topological quantification of the strength of clustering of a model. We investigate the model dependent variation of the number of peaks per island for the 3D Gaussian field models in Section 3.6.3.

3.3.6 Intensity, difference and ratio maps

Here we briefly revisit the concept of intensity maps (see also Equation 2.9). Also, we introduce the difference and ratio maps.

Intensity maps

We divide the persistence diagrams into $n \times n$ regular grid cells. The intensity in a grid cell (i, j) is the total number of dots $N_{i,j}$ in that cell. The intensity map are averaged representations of persistence diagrams over multiple realizations. We define

the normalized intensity function I_{ij} as

$$I_{ij} = \frac{\langle N_{ij} \rangle}{\langle N_{tot} \rangle}, \tag{3.21}$$

where $\langle\rangle$ denotes the ensemble average over many realizations. Here $\langle N_{tot} \rangle$ is the total intensity, or total number of points over $\mathbb{R}^2$, which is simply the total number of topological holes of a particular dimension that form and destroy as the density threshold is lowered. The intensity function $I : \mathbb{R}^2 \to \mathbb{R}$ in the *mean density-persistence* plane is such that its integral over every region $R \subset \mathbb{R}^2$ is equal to unity, i.e.

$$\int_R I_{ij}(\vec{x})\, \mathrm{d}x = 1. \tag{3.22}$$

The plane is defined by the mean-density coordinates on the horizontal axis and persistence on the vertical axis.

In 3D, there are three possible maps, one each for islands (D=0), tunnels (D=1) and voids (D=2). Figure 3.13 presents typical maps for Gaussian random field for power-law power spectrum for different spectral indices.

Difference and Ratio maps

It is possible to highlight the difference between the intensity maps of two different models quantitatively. One possible method to this end is to construct the *difference* and the *ratio* maps. The difference of two functions f and g is a signed difference between their intensities at any location in the mean density–persistence plane, while the ratio function is the ratio of their intensities, such that

$$\Delta_{f,g}(i,j) = I_f(i,j) - I_g(i,j) \tag{3.23}$$

$$\tilde{\Delta}_{f,g}(i,j) = \frac{I_f(i,j)}{I_g(i,j)}. \tag{3.24}$$

The difference and the ratio functions quantify the excess or deficit of topological holes of f with respect to g, in the neighborhood defined by a given mean density and persistence value. Evidently, the difference maps are less sensitive to noise than the ratio maps.

3.3.7 Genus and Euler characteristic

For a connected, orientable surface, the *Genus* [2] is defined as the maximal number of independent simple closed curves that can be drawn on the surface without render-

[2]For consistency, it is important to note that the definition of genus g used in cosmological studies is different from the mathematical definition of genus G, the difference being the number of connected regions: $g = G - c$.

ing it disconnected.

The Euler characteristic, sometimes also known as the Euler-Poincaré characteristic (Edelsbrunner & Harer 2010), of a manifold is the alternating sum of simplex numbers of all the dimensions needed to triangulate it. It is best understood through the Poincaré formula. Assume a manifold $\mathbb{M}$. Then the *Poincaré formula* states that

$$\chi(\mathbb{M}) = V - E + F, \tag{3.25}$$

where, V, E and F are respectively the number of vertices edges and faces used to triangulate the manifold. The genus of a surface is closely related to its Euler characteristic. Let us consider a manifold $\mathbb{M}$, and denote the closed surface that bounds it by $\partial\mathbb{M}$. If the genus of $\partial\mathbb{M}$ is g, the Euler characteristic of $\mathbb{M}$ is given by

$$\chi(\mathbb{M}) = 2 - 2g(\partial\mathbb{M}). \tag{3.26}$$

3.4 Persistence characterization of 1D Gaussian random functions

For obtaining insight into the topological characteristics of Gaussian random fields, we first assess 1D Gaussian random functions. We begin by describing the features of the distribution of critical points in the density field. Subsequently we relate it to their topological properties, as described through persistence. The analysis of 1D field is particularly useful for a visual appreciation of how the critical point characteristics are related to the properties of the persistence diagrams. This becomes an important exercise in view of the observation that a visualization of the interior of the structures in 3D fields is a considerably challenging task, due to complications involved in rendering and visualizing the interiors of 3D objects.

The features visible in the function can be characterized in terms of their mean density, persistence and scale. Recall that every topological feature is associated with two critical points. In the 1D case, there is a local maximum responsible for its birth, and a local minimum responsible for its death. The *mean density* of a feature is half the sum of absolute values of the maximum and the minimum. The *persistence* of a feature is defined as the absolute difference between the maximum and the minimum. The *scale* is defined as the separation between the two adjacent maxima or minima. To a first approximation, this can be deemed roughly equal to the scale on which the field is smoothed.

Figure 3.3 presents the density fluctuations for a random realization of a 1D Gaussian random function. We plot the spatial coordinate $\mathbf{x}$ on the horizontal axis, and the density $\delta(\mathbf{x})$ at that location on the vertical axis. The function values are normalized by the rms of the density fluctuations. The function is characterized by a power law power spectrum. The normalized power spectrum $P(k)$ is presented in the top-left panel of Figure 3.1. The bottom-left panel of the same figure presents the power per logarithmic bin, $k^3\,P(k)$. It is indicative of the relative variance of the models with respect to each other. From the Figure, it is clear that the variance of the field decreases with decreasing spectral index.

The $n = 1$ model is characterized by increasing power at smaller scales. Hence we see a dominating presence of small scale features with high amplitude in the den-

sity field. In terms of critical points, this denotes a presence of spatially close high peaks separated by low valleys, giving the field a very spiky appearance. As the index of the power spectrum decreases, there is increasing power at larger scales, accompanied by diminishing power at smaller scales. This results in large scale features becoming more prominent in the density field. Low power at small scales manifests as tiny wiggles modulating the prominent large scale features.

The features of the density field in Figure 3.3 are reflected directly in the persistence diagrams shown in Figure 3.9. The diagrams are symmetric about $\nu = 0$, reflecting the symmetry about the mean of the field. Maximal persistence decreases marginally for decreasing spectral index. This can be tied in to the fact that the variance of the fluctuations in the box decreases with decreasing spectral index. Indeed, for a given filtering scale R_f, the variance of the density fluctuation goes as (Bardeen et al. 1986)

$$\sigma_0(R_f) \propto R_f^{-(3+n)}. \tag{3.27}$$

The diminishing variance, as a function of the decreasing power spectrum, is responsible for the decrease in the separation between the pair of maximum and the minimum that constitutes the maximal-persistence feature in the different models. This trend can be noticed in the persistence diagrams, where we see the maximum persistence falling monotonically with decreasing spectral index. The fact that the height of the maximal peaks decreases, with decreasing spectral index, can also be confirmed from the structure of the 1D function in Figure 3.3. It can also be noticed that as the spectral index decreases, the concentration of dots increases towards the horizontal axis. This is a direct reflection of the increase in the number of small scale features of low persistence, as the spectral index decreases.

3.4.1 Cumulative and marginal distributions

The persistence diagrams can also be characterized through the cumulative and marginal distributions of the mean density and persistence of the topological features. The marginal density $n(\nu)$ is defined as the number density of the topological holes in the threshold range $(\nu, \nu + d\nu)$. The normalized cumulative distribution $\mathbb{N}(\nu)$ is defined as

$$\mathbb{N}(\nu) = \frac{\int_0^\nu \mathrm{n}(\nu')\, d\nu'}{\int_0^\infty \mathrm{n}(\nu')\, d\nu'}, \tag{3.28}$$

where, $\mathrm{n}(\nu')$ is the marginal number density between $(\nu', \nu' + d\nu')$. The cumulative distribution asymptotically approaches unity, as $\nu \to \infty$.

Distribution of mean density

Figure 3.10 presents the marginal and the cumulative distribution of mean density of the features for the various models. The left panel plots the marginal distribution of mean density. The curves are symmetric about $\nu = 0$, where they also attain their maximum. This indicates that for a Gaussian field, the constituent maximum and

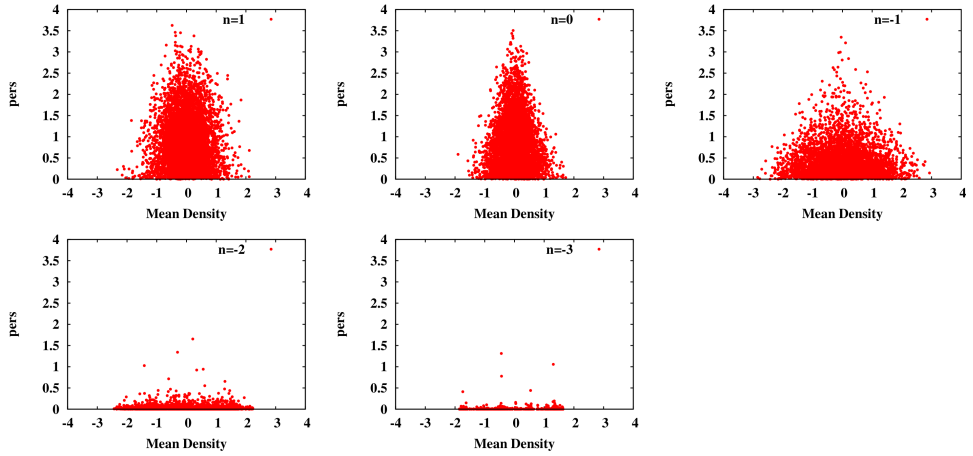


Figure 3.9 Persistence diagrams for the features in the models 1D random functions with power-law power spectra The mean density is plotted on the horizontal axis, and the persistence is plotted on the vertical axis.

minimum for a majority of the features are a mirror image about the mean of the density field. Thus, these are the *zero-mean* features. As the spectral index decreases, the amplitude of the curves falls monotonically. This is simultaneously accompanied by increasingly fatter tails. This indicates an increasing population of features with a non-zero mean density. This phenomenon is intimately connected to the shape of the power spectrum. As the spectral index decreases, the location of the maximum of the curve of the marginal number density of peaks shifts towards lower density thresholds (see Figure 2, Bardeen et al. (1986)). This results in increasingly significant number of peaks for lower density thresholds. In fact, for $n \sim -3$, the amplitude of the curve of the number density of maxima is symmetric about $\nu = 0$, and there are a significant number of peaks even below the mean of the field (Bardeen et al. 1986). The shift of peaks towards lower density thresholds results in features whose constituent pair of maximum and minimum both have function values lower than the mean of the field. As a result, their mean density is lower than the mean of the density field. In fact, it is only required that the maxima-minima pair of a feature be not symmetric about the mean of the field, for it to have a non-zero mean density.

The negative mean density of the features is indicative of the *cloud-in-void* phenomenon (Sheth & van de Weygaert 2004). This happens when a local maximum is located in an overall underdense region. For Gaussian fields, the cloud-in-void phenomenon increases with decreasing spectral index. An important point to remember is that as Gaussian fields are symmetric, the opposite *void-in-cloud* phenomenon also occurs. This means that a number of minima have their function values higher than the global mean. For Gaussian fields, this phenomenon also increases as the spectral index decreases.

The right panel of Figure 3.10 presents the curves for the cumulative distribution

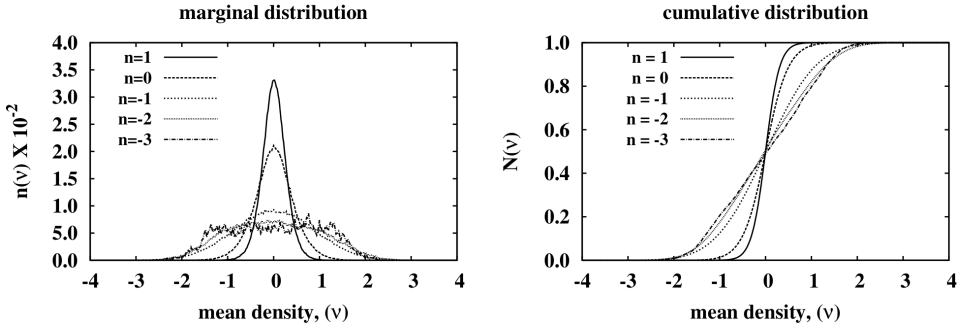


Figure 3.10 Normalized marginal and cumulative distribution of mean density of the topological features. The graphs are drawn for varying spectral index. Left: marginal distribution of the mean density of the features. Right: cumulative distribution of mean density of the features.

of the various models. The steepness of the distribution decreases for decreasing spectral index. The mean density of the major fraction of features for the $n = 1$ model is located within $|\nu| \leq 1$. For lower spectral indices there is a significant fraction of features beyond $|\nu| \geq 1$. The fraction increases with decreasing spectral index, and we see a a progressing uniformity as the spectral index decreases.

Distribution of persistence

Figure 3.11 presents the marginal and the cumulative distribution of persistence of the features. The distribution of persistence π is different from the distribution of mean density ν. The cumulative number density for the $n = -3$ model increases most sharply, followed by $n = -2, -1, 0$ and 1 in that sequence. For the $n = 1$ model, the features are more uniformly distributed along the persistence axis. The smoothly increasing cumulative distribution curve in the right panel of Figure 3.11 confirm this impression. This is indeed what we observe in the persistence diagrams of the 1D Gaussian process in the Figure 3.9. For lower spectral indices, the curve quickly converges to its asymptotic limit 1. It is a clear manifestation of the fact that for lower spectral indices, there is a larger fraction of small scale low amplitude features, and a relatively lower number of large scale features with large amplitude (also see Figure 3.3). The left panel of the Figure 3.11 presents the marginal distribution of persistence for the models. The curves for $n = 1$ and $n = 0$ indicate a clear peak and have a wide tail. The peak disappears for index $n \leq -1$. In this, it resembles the Gamma distribution. The Gamma distribution is the continuous version of the discrete Poisson distribution.

3.5 Persistence characteristics of 3D Gaussian random fields

Having assessed the topological characteristics of 1D Gaussian random functions, we turn to the analysis of the topology of the 3D fields. We present the persistence characterization of the 3D models described in Section 3.2. Subsequently, we inves-

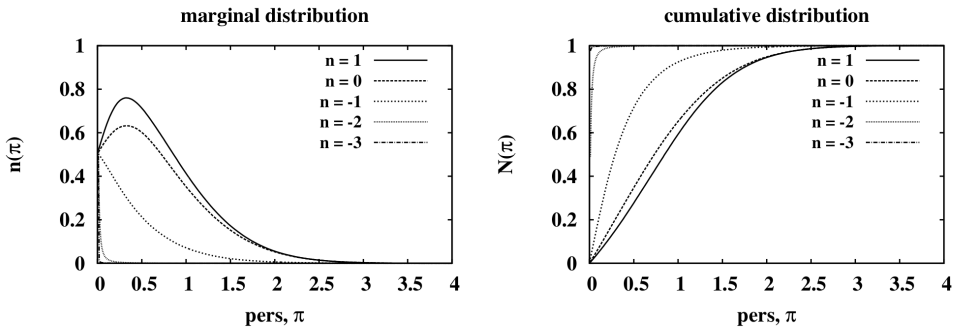


Figure 3.11 Normalized cumulative and marginal distribution of persistence of the topological features. The graphs are drawn for varying spectral index. Left: marginal distribution of persistence of the features. Right: cumulative distribution of persistence of the features.

tigate the topological characteristics of the intensity, difference and ratio maps of the models. We also investigate the properties of the distribution of topological objects as a function of their mean density and persistence.

3.5.1 Number density per unit volume of the topological features

The average total intensity $\langle N_{tot} \rangle$, per unit volume, is a measure of the topological structure of a field. It corresponds to the average number of topological objects per unit volume in the density field. Figure 3.12 presents the average total intensity as a function of the index n of the power spectrum. The average total intensity shows a characteristic dependence on the choice of the power spectrum. It decreases monotonically with decreasing spectral index. This means the number of topological objects per unit volume decreases monotonically with decreasing spectral index. This is expected on account of the fact that for lower spectral indices, the structures in the field become progressively larger. As a consequence, lesser number of features can be packed in a given volume.

3.5.2 Intensity maps

Figure 3.13 presents the intensity maps for the 3D Gaussian random field models. The intensity maps for the LCDM power spectrum is presented in Figure 3.14. The intensity along successive contours decreases by a factor of 2. The left column of the figures presents the intensity maps for islands, the middle column presents the maps for tunnels, and the right column presents the maps for voids.

For all the models, the intensity maps are triangular in shape for all the dimensions, i.e. for the islands, tunnels and voids. The maps also exhibit a symmetry under reflection along the vertical axis: the maps for islands is a mirror image of the map for voids. The map for tunnels is a reflection of itself, the axis of symmetry being $\nu = 0$. This reflects the symmetry of the Gaussian field itself (which has half the volume with negative field values (underdense), and the other half with positive field values (overdense)). The shape of the maps also show various degrees of con-

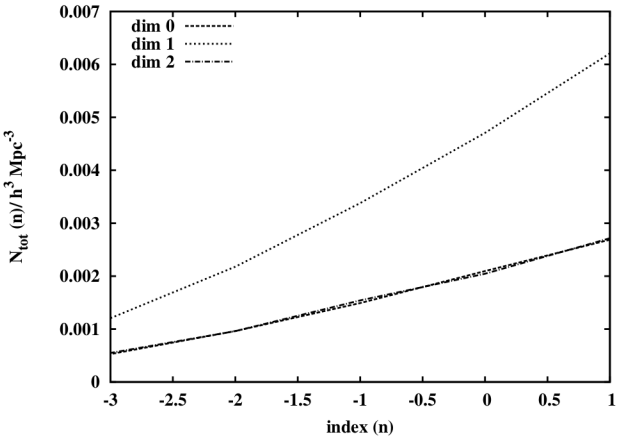


Figure 3.12 The average total intensity as a function of the index of the power spectrum. It indicates the average number of topological objects per unit volume in the density field characterized by a given power spectrum. The curve is presented as a function of the spectral index n. The average total intensity decreases monotonically with decreasing spectral index.

cavity in their arms. The level of concavity depends on the spectral index. For the 0- and 2-dimensional maps, the concavity is in the arm towards which the maps tilt. The quantification of the model dependent concavity appears non-trivial, and may be an interesting exercise for the future.

By definition, the integral of the intensity function over the plane defined by the mean density and persistence is unity. Recasting the intensity function in this normalized format means that the total intensity is conserved, and the same, irrespective of the model. Seen in the light of this statement, the effect of the choice of the model is restricted to the redistribution of intensity in the plane. This has repercussions on the shape of the intensity function, which can be observed by following the iso-contours of a particular intensity in the intensity maps. As an example, we follow a randomly picked iso-intensity contour $I = 0.006$ in Figure 3.15. The peak height of the contour decreases while simultaneously accompanied by a spread along horizontal axis. The increasing intensity near the horizontal axis for lower spectral indices indicates the growing population of low persistence small scale features, due to decreasing power at those scales. As the index of the power spectrum decreases, the sharp cuspy appearance of the contours progresses towards a more smooth peak. We note that iso-intensity contours of any other value follow a similar pattern.

Interesting to note is that the iso-intensity contour for the white noise case (n = 0) departs from the trend followed by the rest of the models. The curve has a more peaky appearance and thinner tails compared to the others. The starkly different behaviour of the white-noise model in comparison with the other models requires requires a careful consideration, and is still under investigation.

The width in the mean density distribution, as well as maximum persistence are important markers of difference in the map characteristics. They are a reflection

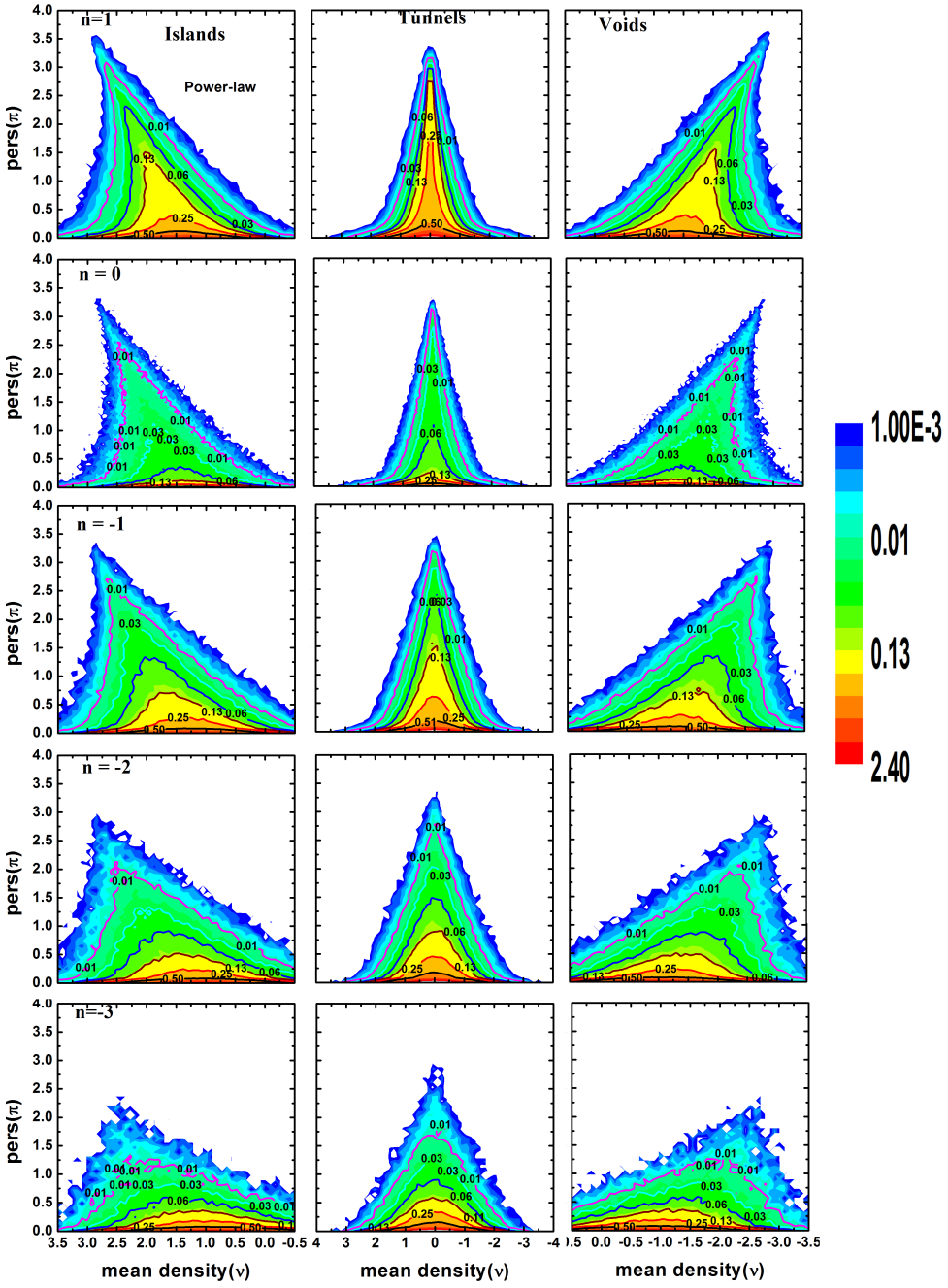


Figure 3.13 Intensity maps of islands, tunnels and voids (left to right) of the 3D Gaussian random field models. Intensity I is defined as the fraction of total objects (for each dimension separately) in a grid cell. Successive contours have intensity values that differ by powers of 2. The maps show a characteristic dependence on the choice of the power spectrum.

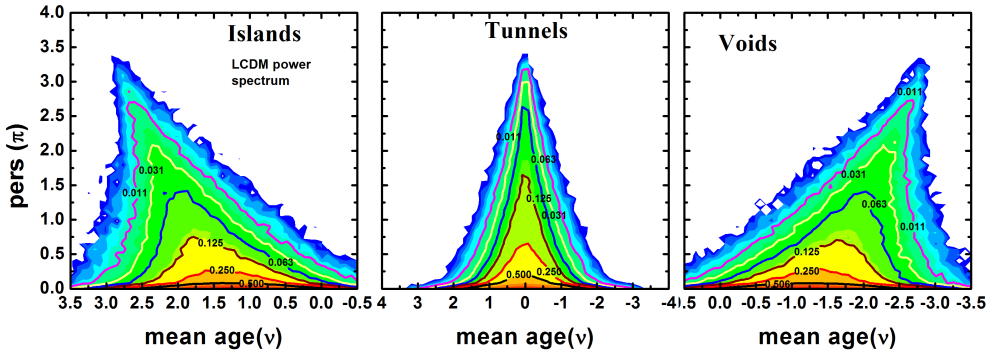


Figure 3.14 Intensity maps of islands, tunnels and voids (from left to right) in Gaussian random fields for the LCDM power spectrum. The box size is $128h^{-1}$Mpc, with a smoothing radius of $2h^{-1}$Mpc. Intensity I is defined as the fraction of total objects (for each dimension separately) in a grid cell. Successive contours have intensity value that differ by powers of 2. The color bars are the same as Figure 3.13.

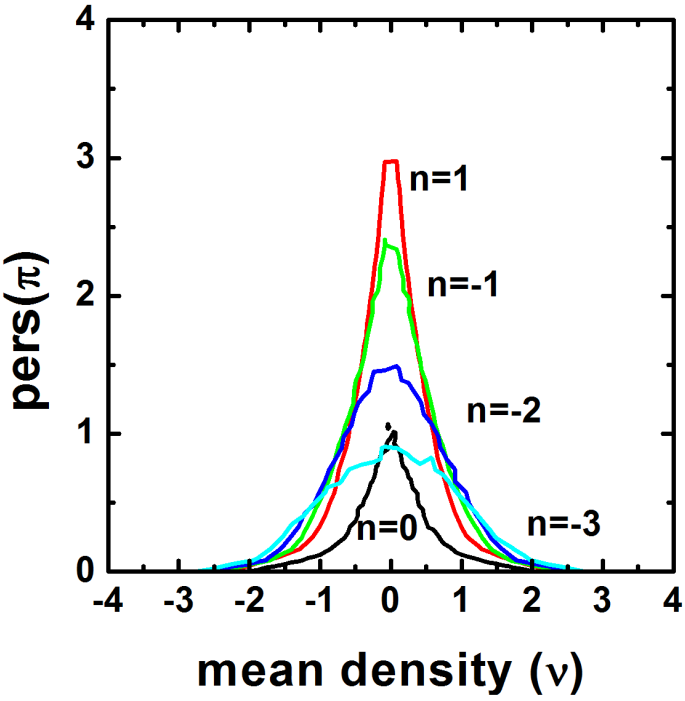


Figure 3.15 Iso-intensity contours corresponding to a randomly chosen value of intensity function, $I_{ij} = 0.06$. The curves are presented for the power-law power spectrum. The curve characteristics show a systematic dependence on the choice of the power spectrum. The curve for white noise departs from the trend followed by curves for other spectral indices.

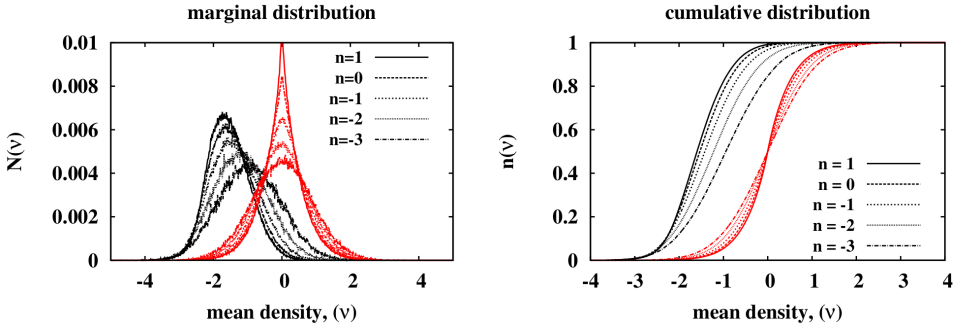


Figure 3.16 Normalized marginal and cumulative distribution of mean density the features in the models of 3D Gaussian fields. The graphs are drawn for varying spectral index. Black: topological voids, red: tunnels. Curves for the topological islands are identical to the curves for the islands, under reflection about $\nu = 0$.

of the features in the density landscape of the field. From $n = 1$ to $n = -3$, the maximum persistence decreases. As argued in the 1D case, this is a reflection of the decreasing variance of the density field, as the spectral index decreases: $\sigma_0 \propto R^{-(n+3)}$.

3.5.3 Mean density: Statistical distribution

In this section, we study the marginal and cumulative distribution of mean density of the topological features. We are also interested in an empirical fitting formula for the distributions. The form of the fitting function is motivated by the observed curve characteristics.

Figure 3.16 presents the marginal and cumulative distribution of the mean density of the features for the various models. The left and the right panels present the marginal and the cumulative distribution respectively. The curves in black present the graphs for the topological voids. The curves in red present the graph for tunnels. The curves for the topological islands are identical to the curves for the voids, under reflection about $\nu = 0$.

The rate of change in the cumulative distributions of voids and tunnels is the same for any given power spectrum. This follows from the observation that the local slope of the cumulative distribution is the same for both voids and tunnels. This may also be visually confirmed from the graphs. The cumulative distribution increases most steeply for the $n = 1$ model, and decreases with decreasing spectral index. The curves for the tunnels cross each other at $\nu = 0$. A reflection of this can also be seen in the curves for the marginal distribution of mean density. The curves for the tunnels are symmetric about $\nu = 0$. The peaks of the curves are located at $\nu = 0$, irrespective of the model.

The cumulative and the marginal distribution curves for the voids are shifted towards lower density thresholds for higher spectral indices. The location of the maximum of the marginal distribution of voids follows a similar trend. As noted

earlier, the curve for islands is a mirror image of the curve for voids. This means that for a smaller spectral index, the cumulative and the marginal distribution curves are shifted towards lower density thresholds. As noted in the 1D case, this is a consequence of the fact that for lower spectral indices, there is a significant fraction of maxima located near or even below the mean of the field.

Mean density fit: the skew-normal distribution

Under the assumption that the intensity maps arise from a well behaved probability distribution function (Section 3.3.6), we attempt an empirical fit to the distribution. To this end, we take into account that the distributions for islands and voids are skewed and a mirror image of each other about $\nu = 0$. We also note that the distribution of tunnels is symmetric to itself and exhibits no skewness. The observation that the curves for islands and voids exhibit a skewed distribution, while the curve for tunnels has no skew motivates to introduce a generic class of distribution called the *skew-normal distribution*, as a fit simultaneously for the marginal distribution of mean density of islands, tunnels and voids. The well known normal distribution is emergent from the skew-normal distribution.

The skew-normal distribution is given by(O'Hagan & Leonard (1976); Azzalini (1985)):

$$f(\nu) = \frac{A_0}{\omega\pi} \mathrm{e}^{-\frac{(\nu-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{\nu-\xi}{\omega}\right)} \mathrm{e}^{-\frac{t^2}{2}}\, \mathrm{d}t, \tag{3.29}$$

where, A is the amplitude, α is the *skewness parameter*, ξ is the location parameter and ω is the scale parameter. For more details on the skew-normal distribution, refer to Appendix G.

We present the fit to the marginal distribution curves for the $n = 0$ model in Figure 3.17. The solid red curves present the marginal distribution of islands, tunnels and voids for the $n = 0$ model. The black dot-dashed lines presents the skew-normal fit curves for the same. In addition, the curves for $n = 1$ and $n = -1$ are also presented in red, for comparison as dotted and dot-dashed lines in red respectively. The fitted curves for islands and voids match the actual distribution remarkably well. The curve of actual distribution for tunnels has slightly broader tails and higher peak than the fitted curve. The fit for tunnels has the skewness factor $\alpha = 0$, indicating that the mean density of the tunnels may be normally distributed.

Table 3.1 presents the parameters of fit for the various power-law models using the skew-normal distribution. Columns 3 presents the skewness, defined $\gamma_1 = \mu^3/\sigma^3$. Column 4 presents the excess kurtosis, defined as $\gamma_2 = \frac{\mu^4}{\sigma^4} - 3$. By this definition, the excess kurtosis of the standard normal distribution is 0. The absolute value of skewness for the curves corresponding to islands and voids increases as one progresses from $n = 1$ to $n = -3$. The distribution for tunnels exhibits a near-zero, negligible skewness for all values of n. The excess kurtosis for all islands, tunnels and voids is $\sim -2.9 : -3.0$, with small variations that show a decreasing trend in the absolute magnitude as one lowers the spectral index.

	index (n)	Relative intensity ($\Lambda_{f1,f2}$)	Skewness (γ_1)	Excess Kurtosis (γ_2)	*Mean-density* Amplitude (A_0)	*Mean-density* Location param.(ζ)	*Mean-density* Skewness param.(α)	*Mean-density* scale/width (ω)	*pers.*(γ) Amplitude (ψ_0)	*pers.*(γ) Dimension (τ)
	1	1.31	0.006	-2.97	908.26	4.30	-1.91	1.73	92.48	0.92
	0	1.00	-0.018	-2.96	693.34	4.22	-1.65	1.83	69.86	0.91
Islands	-1	0.72	-0.064	-2.95	502.95	4.01	-1.38	1.94	48.53	0.91
	-2	0.46	-0.135	-2.93	326.56	3.52	-0.99	1.99	26.79	1.00
	-3	0.24	-0.235	-2.89	181.87	3.11	-0.98	2.25	8.90	1.26
	1	1.31	-0.00072	-2.98	3845.54	0	0	0.95	186.51	0.92
	0	1.00	0.00048	-2.98	3005.86	0	0	1.09	145.87	0.90
Tunnels	-1	0.72	-0.00106	-2.97	2214.19	0	0	1.29	101.99	0.90
	-2	0.47	-0.00099	-2.96	1468.61	0	0	1.52	56.60	1.00
	-3	0.26	-0.00044	-2.94	814.25	0	0	1.75	18.48	1.28
	1	1.31	-0.005	-2.97	910.41	-4.32	1.97	1.75	92.88	0.91
	0	1	0.02	-2.97	696.41	-4.22	1.69	1.83	70.56	0.90
Voids	-1	0.72	0.067	-2.96	502.84	-4.03	1.41	1.95	48.28	0.91
	-2	0.46	0.13	-2.93	327.71	-3.56	1.04	2.01	26.47	1.01
	-3	0.24	0.23	-2.89	182.32	-3.11	1.01	2.29	9.02	1.25

Table 3.1 Statistical parameters of the intensity maps as a function of varying spectral index for the power-law models. The parameters are shown for islands, tunnels and voids. Column 1 shows the value of the spectral index, column 2 presents the relative intensity, $\Lambda_{f1,f2}$ which is calculated with respect to the shot-noise maps (power-law, n=0). Columns 3 & 4 present the shape description parameters for the mean-density histograms – Column 3 presents the skewness factor $\gamma_1 = \mu^3/\sigma^3$, which is an indicator for asymmetry around mean and Column 4 presents the excess kurtosis, $\gamma_2 = \frac{\mu^4}{\sigma^4} - 3$, with respect to the standard normal distribution. Columns 5 through 10 present the parameters of fit for the fitting functions of mean-density(column 5–8) and persistence(columns 9–10) histograms. The mean-density histograms for all islands, tunnels and voids are fitted with a skewed-normal distribution as given by equation 3.29. Peaks and voids are visibly skewed distributions with significant values of skewness parameter α. Loops have a symmetric distribution with $\alpha \sim 0$. The persistence histograms are fitted with a power-law distribution. The value of *persistence-dimension* (τ) is presented in column 10. Indices for $n = 1, 0 \& -1$ show similar values of ~ 0.9, and start increasing significantly thereafter for lower values of spectral index.

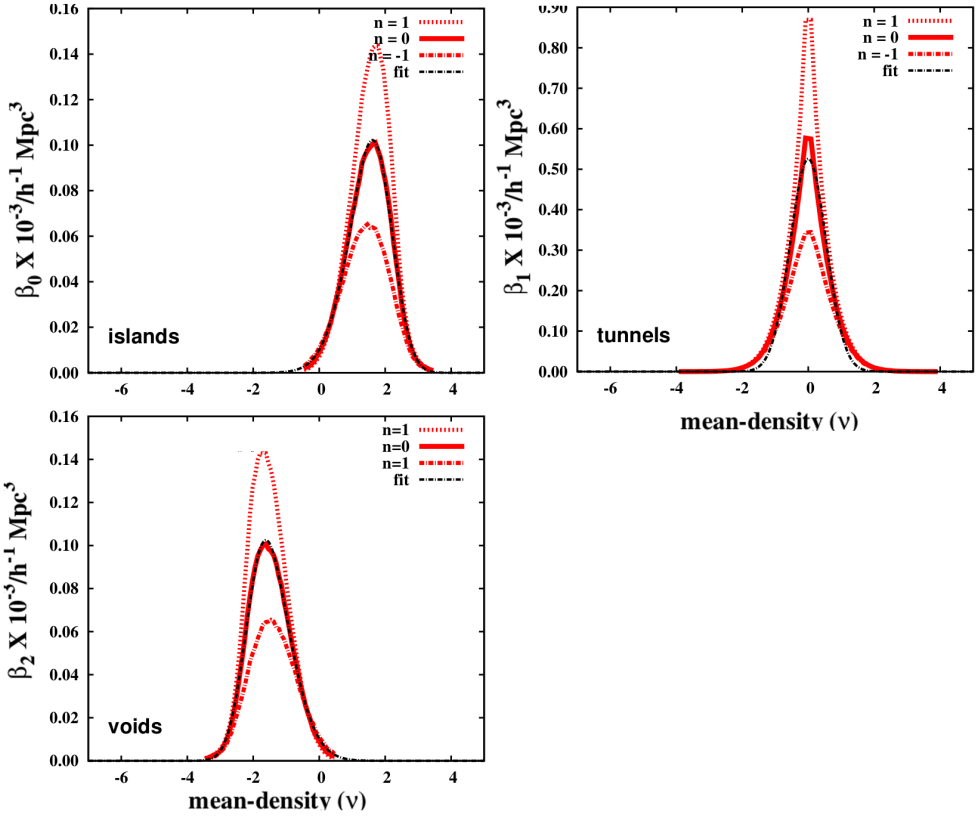


Figure 3.17 Marginal distribution of the mean density of islands, tunnels and voids for $n = 1, 0$ and -1. The solid curve in red is the average computed from multiple realizations. The curve in black(dot-dashed) plots the best fit skew-normal distribution for each of islands, tunnels and voids. The values of parameters of distribution A(amplitude), ζ (location), α (skewness-parameter) and ω(scale/width) are depicted in Table B.1.

The dependence of the various parameters of fit on the value of spectral index is presented in Figure 3.18. The location parameter (ζ) and skewness ($\gamma 1$) show an increasing trend with increasing value of the spectral index, while the width parameter (ω) and skewness parameter(α) show a decreasing trend. Columns 5 through 8 of Table 3.1 also enumerate the best-fit values for the parameters. The absolute value of the location parameter decreases for both islands and voids as n decreases. The trend in the dependence of location parameter on the value of spectral index is depicted by the solid curve in Figure 3.18. The trend in the dependence of skewness parameter α on the choice of spectral index is depicted by the dashed curve in Figure 3.18. The absolute value of skewness parameter decreases for decreasing n in the case of islands and voids. On the other hand, it is ~ 0 for tunnels, as expected. The trend in the dependence of scale/width parameter on the choice of spectral index is depicted by the dot-dashed curve in Figure 3.18. The scale parameter ω, which

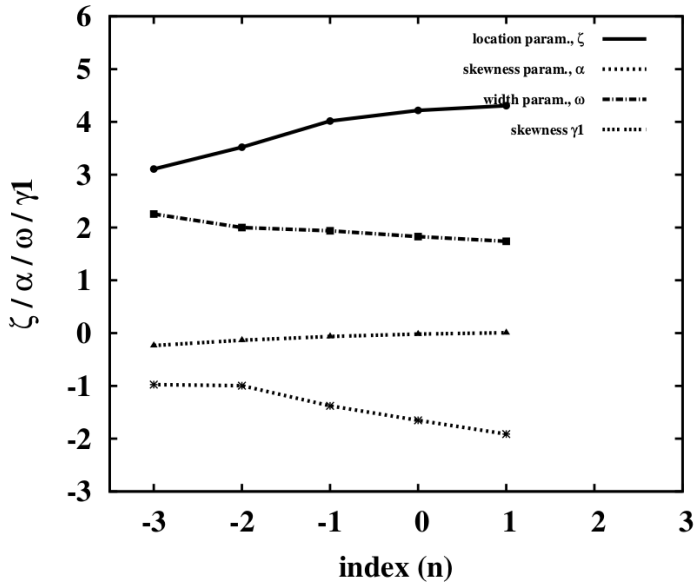


Figure 3.18 Fitting parameters for the mean density distribution. The solid line corresponds to the best-fit values of the location parameter ξ, the dashed line for skewness parameter α, and the dot-dashed line for scale/width parameter ω. The double dashed line corresponds to the skewness (note that skewness is different from skewness parameter).

indicates the width of the curves, increases uniformly for decreasing n. This is true for islands, tunnels and voids.

3.5.4 Persistence: statistical distribution

Figure 3.19 presents the cumulative and marginal distribution of persistence of the features for the various models. The left and the right panels present the marginal and the cumulative distribution. The curves in black present the graphs for the topological voids. The curves in red present the graph for tunnels. The curves for the topological islands are identical to the curves for the voids, due to the symmetry of the field. The marginal distribution of tunnels is different from the marginal distribution of islands and voids. However the cumulative distribution of islands, tunnels and voids are approximately coincident for all the models.

The left panel of Figure 3.20 plots the marginal distributions of persistence for the islands, tunnels and voids for the $n = 0$ model. For low values of persistence $\pi \sim (0.1 - 1\sigma)$, the distribution follows a power-law

$$\psi = \psi_0 \pi^{-\tau}. \tag{3.30}$$

The index of the power-law τ indicates the rate of fall in the number of objects as a function of persistence. We call this the *persistence index*. Columns 9 and 10 of

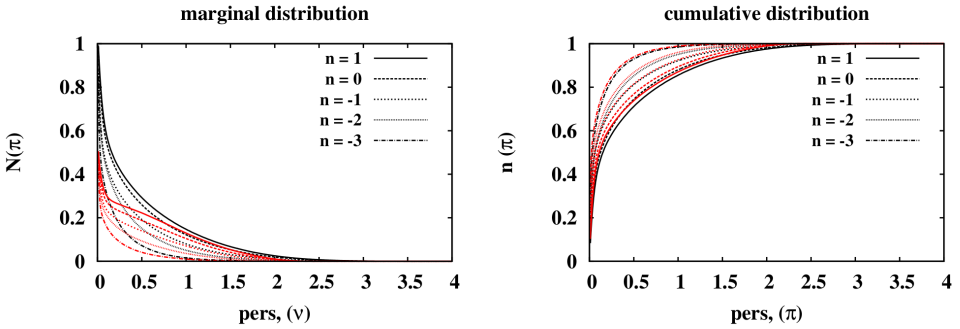


Figure 3.19 Normalized marginal and cumulative distribution of persistence of the features. Black: topological voids, red: tunnels. Curves for the topological islands are identical to the curves for the islands, under reflection about $\nu = 0$.

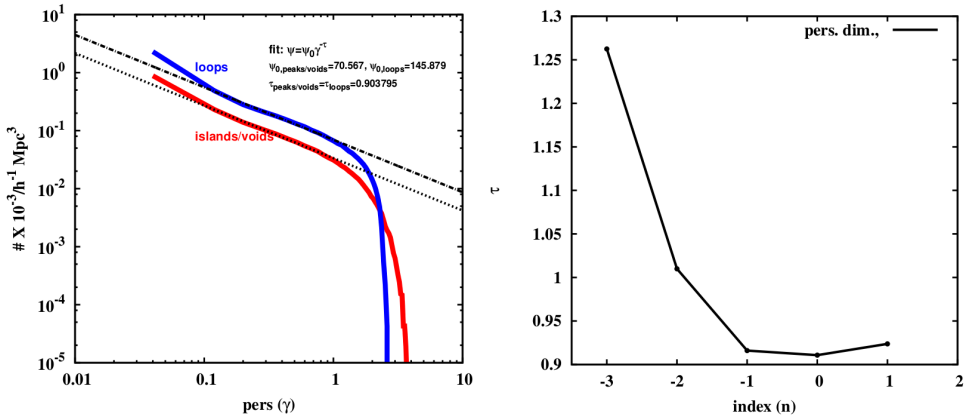


Figure 3.20 Left: Marginal distribution of persistence of the toplogical holes. The rate of decrease is the same for islands tunnels and voids. For small values of persistence, the rate of fall fits a power law. Right : the trend in the dependence of the value of the persistent index on the index of the power spectrum. Persistence index decreases while going from spectral index 1 to 0, and increases monotonically while going from 0 to -3.

table 3.1 list the amplitude ψ_0 and the value of the persistence index τ for the power-law models. The persistence index τ is approximately the same for islands, tunnels and voids for a given spectral index n. For $n = 1, 0, -1$, the persistence index is $\tau \sim 0.9$. It shows an increasing trend for $n < -1$. For the $n = -2$ model, $\tau \sim 1$, and for the $n = -3$ model $\tau \sim \frac{5}{4}$. The right panel of Figure 3.20 presents the trend in the dependence of the persistence index (τ) on the power spectrum index n.

3.5.5 Difference and ratio maps

To compare the intensity maps, we construct the difference and the ratio maps as defined in (3.24). Quantification of the difference and the ratio function $d_{i,j}$ between two models f and g results in maps which are, locally, at various factors of elevation

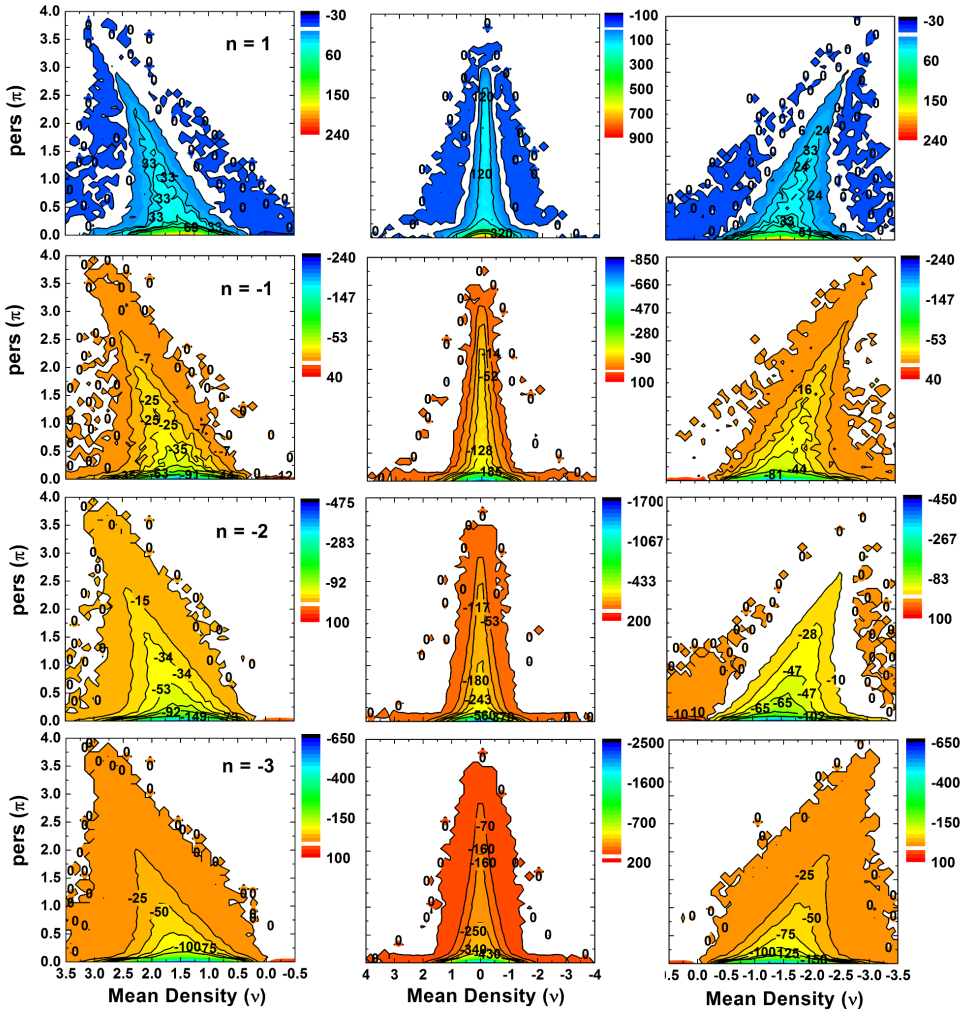


Figure 3.21 Difference maps of islands, tunnels and voids (from left to right) in Gaussian random fields for a power-law power spectrum, for spectral indices $n = 1, -1, -2$and -3. The difference function d is defined as the signed difference of intensities of a given power-spectrum, with respect to the white-noise power spectrum in a grid cell. Successive contours have intensity value as half of the previous ones.

or depression with respect to each other. These elevations and depressions indicate an excess or deficit in the number of topological features in the neighborhood defined by a particular value of mean density and persistence. We note here that the difference and the ratio maps display similar global features. However, the contours in the difference maps are less noisy compared to the ratio maps. This reflects the fact that the difference function numerically is more stable than the ratio function.

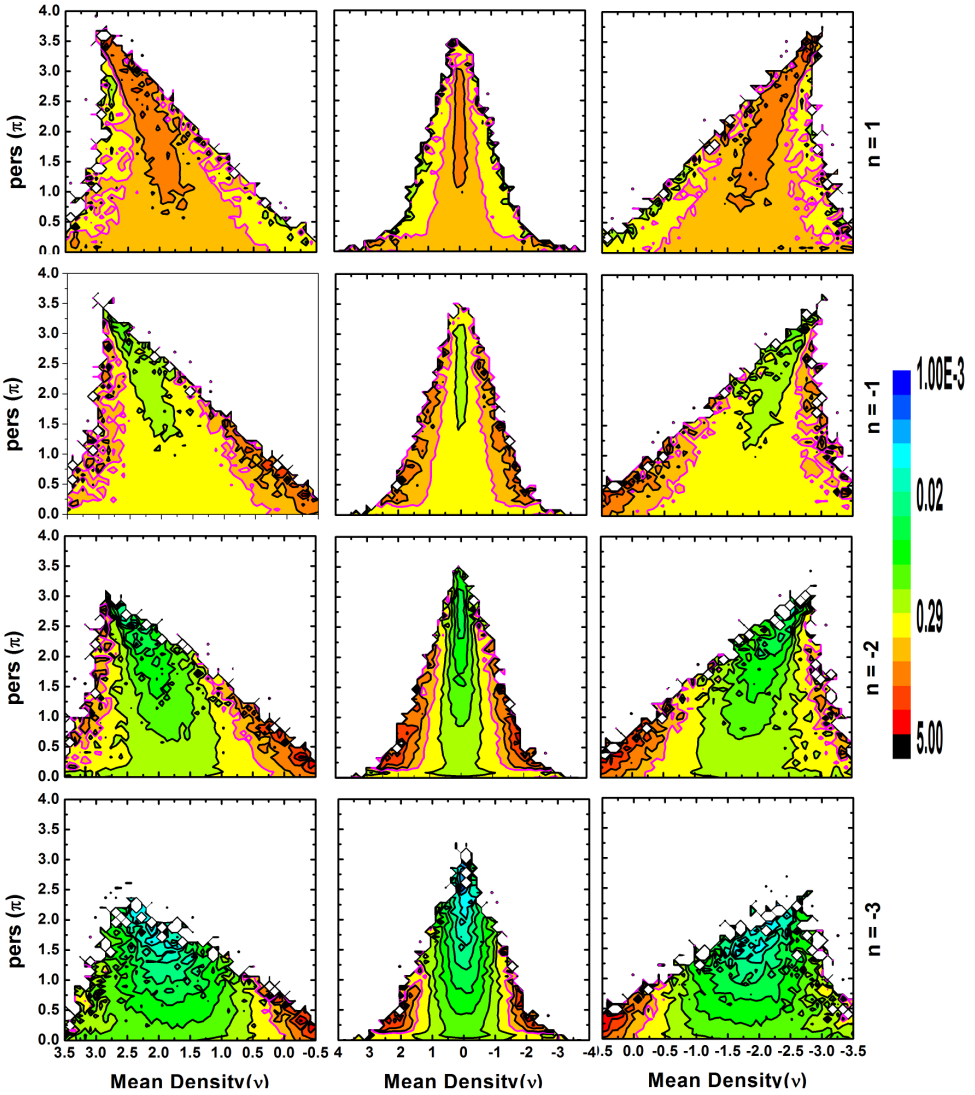


Figure 3.22 Ratio maps of islands, tunnels and voids (from left to right) of Gaussian random fields for a power-law power spectrum, with spectral indices $n = 1, -1, -2$and -3. The maps are constructed with respect to the $n = 0$ model. The box size is $128h^{-1}$Mpc, with a smoothing radius of $2h^{-1}$Mpc. The ratio function Δ is defined as the ratio of intensities of a given power-spectrum, with respect to the white-noise power spectrum in a grid cell. Successive contours have intensity value as half of the previous ones.

Difference and ratio maps of the power law models.

Figure 3.21 presents the difference maps for Gaussian random fields with power-law power spectra. The model of reference for the comparison is the white noise model

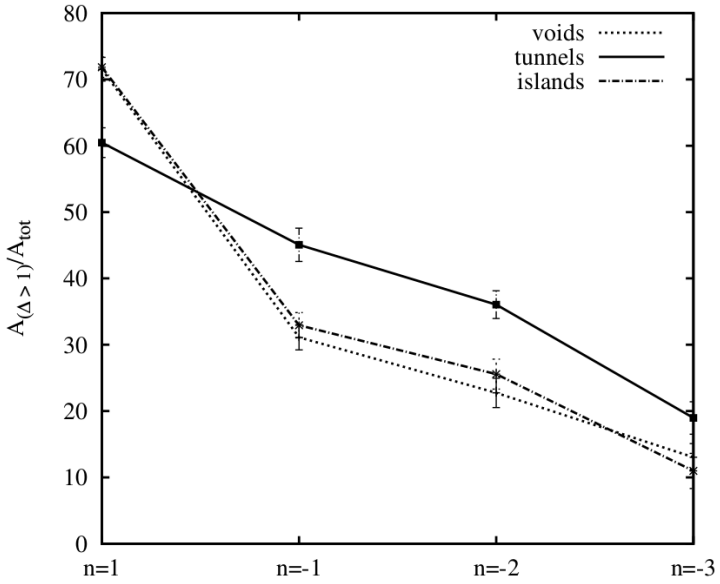


Figure 3.23 Plot of fractional area covered by $\Delta_{ij} > q$ for the different power-law models. The fractional area covered is defined as $A_{\Delta_{ij}>1}/A_{tot}$. For the power-law models, the fractional area covered by the islands and the voids is the same within 1σ error. The fractional area covered for the tunnels is significantly different from those of islands and voids. For the LCDM model, the fractional area covered is the same for islands, tunnels and voids, within 1σ error.

with spectral index $n = 0$. The difference function is given by

$$d_{ij} = I_{(ij),f} - I_{(ij),n=0}, \tag{3.31}$$

such that $I_{(ij),f}$ is the intensity of the compared model in the (i,j)-th bin, and $I_{(ij),n=0}$ is the intensity of the reference $n = 0$ model in the same bin. We follow the maps for tunnels, noting that the maps for islands and voids show similar properties. Iso-difference contours are drawn and marked for a range of values. The details of the features present in the difference maps show a systematic dependence on the value of the spectral index. A visual inspection of Figure 3.21 reveals this. The difference maps, like the intensity maps, also exhibit a symmetry: the maps for islands are symmetric with respect to the maps for voids about $\nu = 0$. The maps for loop is a mirror image of itself under reflection about $\nu = 0$. For all the models, the overall shape of the maps is triangular.

For the maps corresponding to tunnels, the $n = 1$ model exhibits an elevated peak along and around $\nu = 0$. This indicates that the $n = 1$ model has a larger number of topological features with zero or near-zero mean density, as compared to the white noise case. This elevation decreases in height monotonically on either sides. In contrast, the maps for $n = -1, -2$, and -3 are depressed around $\nu = 0$. The extent of depression shows dependence on the value of the spectral index. The

maps corresponding to a lower index are more depressed compared to a map with higher index. This is similar to the trend seen in the 1D case (Section 3.4). As the spectral index decreases, the number of features with the mean density equal to the mean of the field decreases monotonically. It is simultaneously accompanied by an increase in the features with a non-zero mean density. As noted in the 1D case, this happens because lowering the spectral index results in an increase of the small scale features with a non-zero mean density.

The maps show an opposite trend along the edges of the triangular arms, compared to the neighborhood defined by $\nu = 0$. As noted earlier, the $n = 1$ exhibits an excess of features along $\nu = 0$ compared to the white noise case. However, the $n = 1$ model has lesser number of features along the triangular arms compared to the white noise model. This is brought out sharply in the ratio maps for the models, presented in Figure 3.22. The ratio maps maginify the contrast of the differences between the models.

The difference in the map characteristics is also illustrated if one plots the fractional area enclosed within regions defined by a particular ratio value. For a given ratio value q, we define the fractional area as the area bound by the contour q over the total area covered by non-zero regions of the ratio maps

$$A = \frac{\int\int_q dS}{\int\int dS}, \tag{3.32}$$

where, $\int\int_q dS$ is the area above the region demarcated by the contour q, and $\int\int dS$ is the total area of non-zero regions in the ratio maps. Figure 3.23 plots the fractional area covered for $q = 1$, for the different power-law models. Along this contour, the compared models have equal number of features. The fractional area covered by the islands and the voids is the same within 1σ error. The fractional area covered for the tunnels is significantly different from those of islands and voids. The fractional area covered decreases monotonically for all islands, tunnels and voids as the spectral index decreases. The excess of number of features, as denoted by a non-zero value of fractional area for $q = 1$, indicate that all the models have a number of features in excess of the the white noise model for some values of mean density and persistence. This excess in the number of features arises because every model has excess power compared to the white noise model, for some scale.

Difference maps of the LCDM model

Figures 3.24 presents the difference maps of the LCDM model with respect to the power-law models. The left column in the figure presents the maps for islands, the middle column presents the maps for tunnels, and the right column presents the maps for voids.

We are motivated to compare the LCDM model with all the power law models. This is because the LCDM model has a running, scale-dependent spectral index, and its slope locally resembles a power law, with the spectral index $n_{eff} = d\,ln\,P(k)/d\,ln\,k$. Recall from Figure 3.2 that the effective spectral index n_eff for the LCDM model runs between approximately -0.5 and -2.5 in our simulation box.

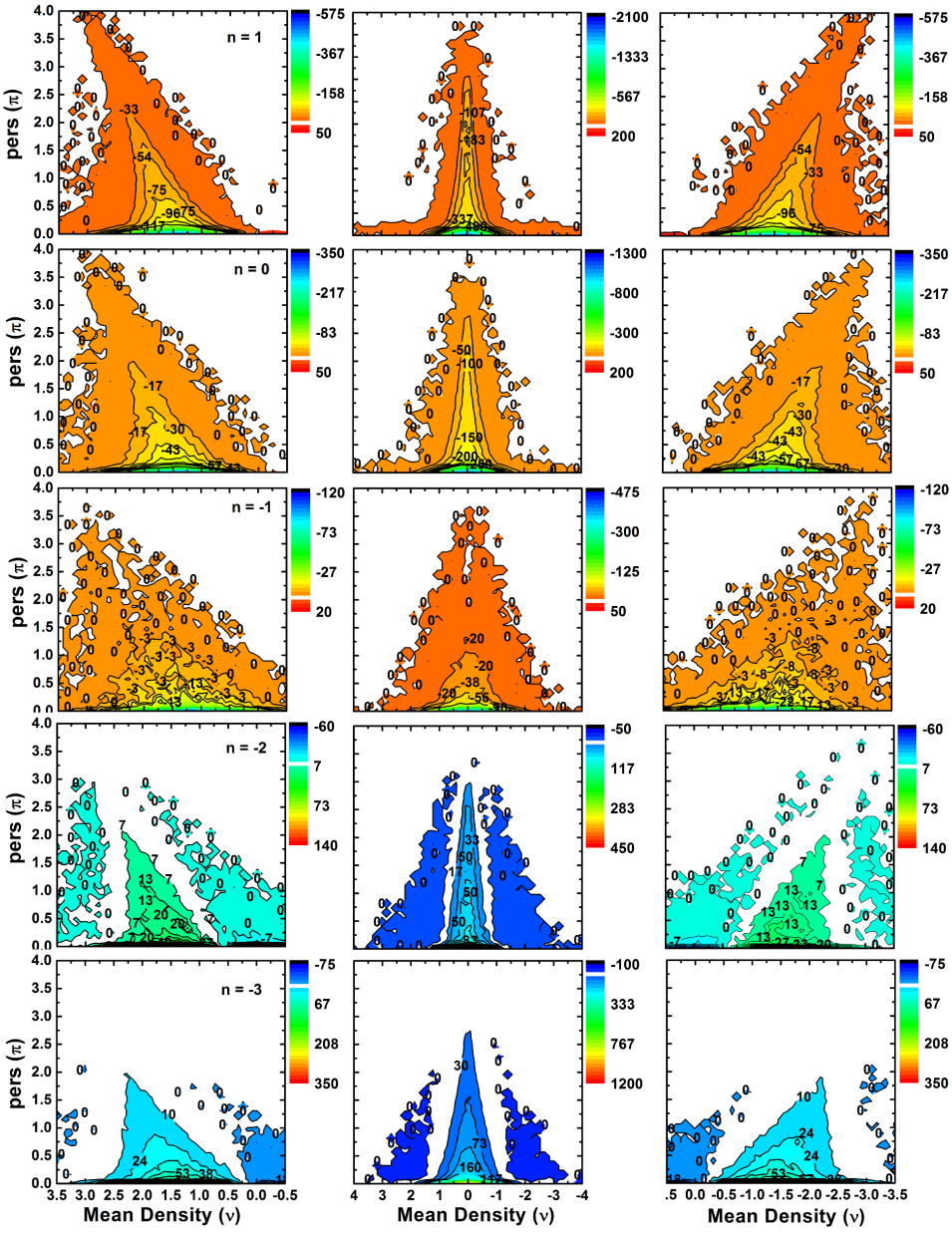


Figure 3.24 Difference maps of the LCDM model with respect to the various power-law models.

The difference maps of the LCDM model show evident differences with respect to the power-law models. The region around the mean density $\nu = 0$ is the region where most of the significant differences arise across all the maps. Examining Figure 3.24, we notice that the the LCDM model has consistently lower number of topological features globally with respect to the $n = 1, 0$ and -1 power-law models. For $n = -2$ and -3, the LCDM model has a higher number of features than the power law models. The $n = -1$ model appears closest to the LCDM model. As can be verified from the values of the contours in Figure 3.24, the difference function reaches a minimum for the $n = -1$ model and the LCDM model.

3.6 Betti numbers of 3D Gaussian random fields

Following the hierarchical topology of the Gaussian field models in terms of persistence, we turn to the analysis in terms of the Betti numbers in this section. We compare them with the topology as described through the Euler characteristic. The three Betti numbers β_0, β_1 and β_2 are computed as a function of the dimensionless density threshold value $\nu = \delta/\sigma$. We also devote a subsection to quantifying the statistics of local maxima or peaks (Bardeen et al. 1986), and the isolated connected objects or islands as quantified by the zeroth Betti number β_0. This is motivated by the need to understand the relation between peaks and islands, and how a topology description in terms of one compares to the other.

In a related article, we present a preliminary investigation of Betti numbers of Gaussian fields, focusing on a comparison with genus statistics (Park et al. 2013). Here we expand and deepen this analysis. In this context, we also note that a semi-analytical calculation of the Betti numbers of the Gaussian fields in 2D has been carried out by Feldbrugge et al. (2015).

3.6.1 Gaussian Betti numbers in 3D: generic properties

The top-left panel of Figure 3.26 shows the unscaled Betti number curves for the various power-law models, as well as the LCDM model. For Gaussian fields, β_0 and β_2 are a mirror image of each other about $\nu = 0$. β_1 is symmetric to itself under reflection about $\nu = 0$. Because of their symmetry, an analysis with respect to the islands is also indicative of the properties of voids.

At $\nu =\sim \pm\sqrt{3}$, the number of isolated islands and voids attain their maximum respectively. The maximum of β_0 curve gradually moves to lower thresholds as n decreases. At $\nu \sim 1$, the number of isolated islands is equal to the number of isolated tunnels. This threshold of equality gets closer to the maximum of β_0 as the spectral index decreases. At $\nu = 0$, the number of isolated tunnels attains a maximum. At this threshold, the number of isolated islands also equals the number isolated voids. The extent of overlap between islands and voids increases with decreasing spectral index. It is negligible for the $n = 1$ model, and becomes considerably significant as the spectral index decreases to $n = -3$.

Figure 3.25 illustrates the iso-density surfaces denoting the structure of the field for three different density thresholds $\nu = \sqrt{3}, 1,$ and 0, for the $n = 1$ and the $n = -3$ models. The left column presents the contour surfaces for the $n = 1$ model, and the

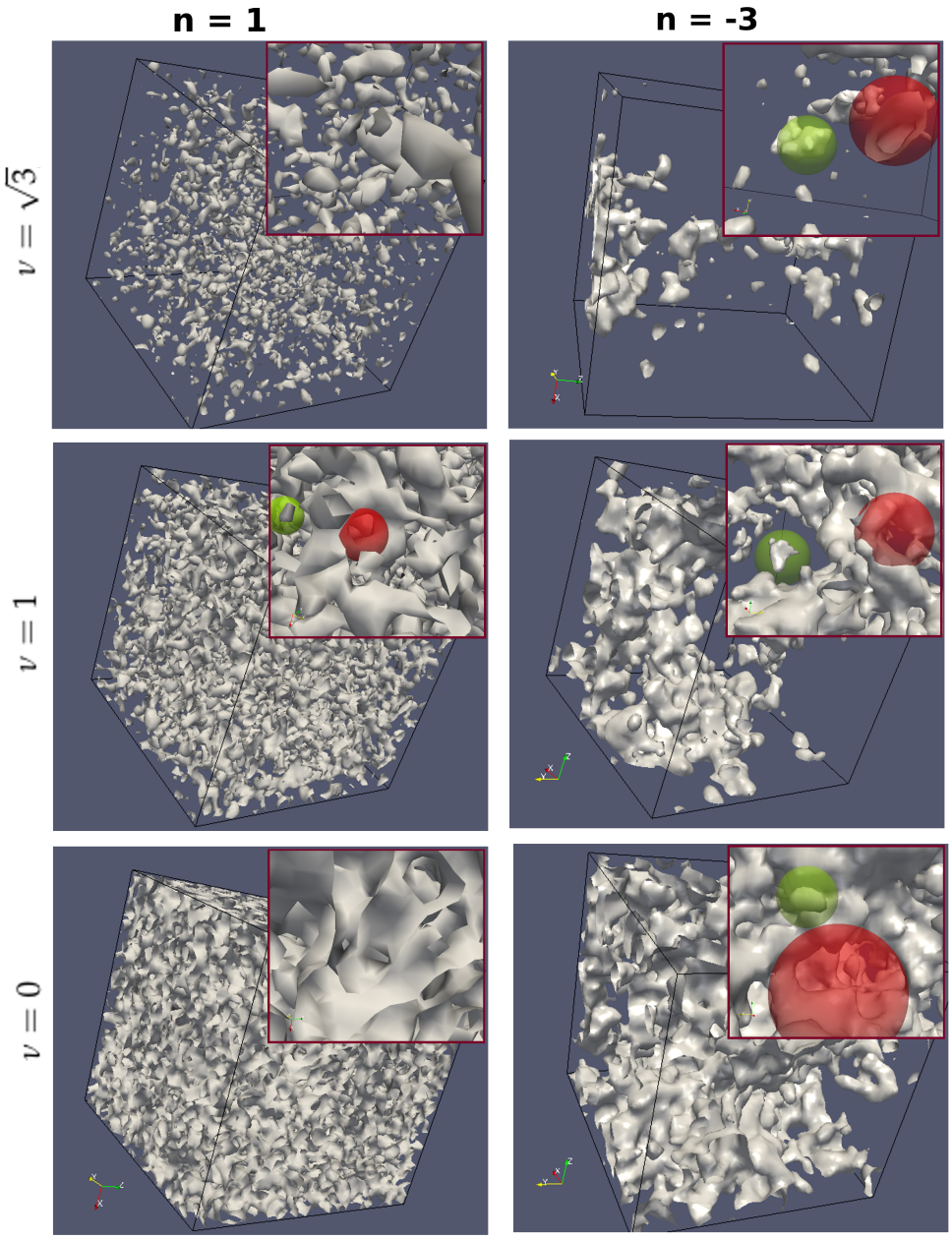


Figure 3.25 Caption next page

Figure 3.25 Iso-density surfaces denoting the structure of the field for three different density thresholds $\nu = \sqrt{3}, 1$, and 0, for the $n = 1$ and the $n = -3$ models. The left column presents the iso-density surfaces for the $n = 1$ model and the right column presents the contour surfaces for the $n = -3$ model. Examples of typical tunnels are enclosed in translucent red spheres; examples of typical isolated islands are enclosed in green spheres. The topology of the contour surfaces shows a dependence on the choice of the power spectrum, as well as the density threshold.

right column presents the contour surfaces for the $n = -3$ model. For illustration, examples of a typical tunnel are enclosed in a translucent red sphere. Examples of a typical isolated island are enclosed in a green sphere.

The topology of the iso-density surfaces shows a dependence on the choice of the power spectrum. The dependence on the power spectrum is most evident around $\nu = 0$. At this threshold, for the $n = 1$ model, the topology is predominantly loop-like or sponge-like, with nearly a singly connected surface. For the $n = -3$ model, the topology is a visible mixture of loops as well as isolated islands. The strength of mixing of the different kinds of topological entities, indicated by the relative number of different topological holes, increases with decreasing spectral index. This indicates that the different models have inherently different topological structure. In other words, the manifold cannot be described as either predominantly meatball-like or sponge-like, but rather as a mixture of these topologies.

The increase in the overlap between meatball-like, sponge-like and cheese-like topologies for decreasing spectral index is more evident once we zoom in to the concerned region in Figure 3.27. The bottom-left panel plots the overlap between β_0 and β_1, and the bottom-right panel plots the overlap between β_0 and β_2. The value at the point of overlap increases with decreasing spectral index for both β_0/β_2 as well as β_0/β_1. These trends are related to the observation that for lower spectral indices there is significant power at large scales. This is reflected in the presence of a larger number of isolated islands, even significantly below the mean value of the density field. This is indicative of the *cloud-in-void* effect (Sheth & van de Weygaert 2004). The number of *clouds* or islands in the overall underdense regions increases with decreasing spectral index. The opposite *void-in-cloud* effect also increases with decreasing spectral index. This means that there are more and more void like regions above the mean of the density field, as the spectral index decreases.

The presence of a progressively larger number of isolated islands for low rms density thresholds, or the increasing cloud-in-void effect, is also evident if one plots the fractional value of number of islands with respect to the sum of total number of islands, tunnels and voids. This is shown in the left panel of Figure 3.28, where we plot this quantity as a function of the dimensionless density threshold. For higher spectral indices, the fractional value quickly falls down to almost zero. On the other hand, for lower spectral indices it is larger than zero for thresholds as low as -2σ.

The shape of the Betti number curves show a dependence on the choice of the power spectrum: the Betti number curves become broader as n decreases. This can be seen from the shape of the scaled Betti number curves in the top-right panel of Figure 3.26. The Betti numbers have been scaled with respect to the $n = 0$ model. It is also evident from the overlap between the various Betti numbers (see Figure 3.27).

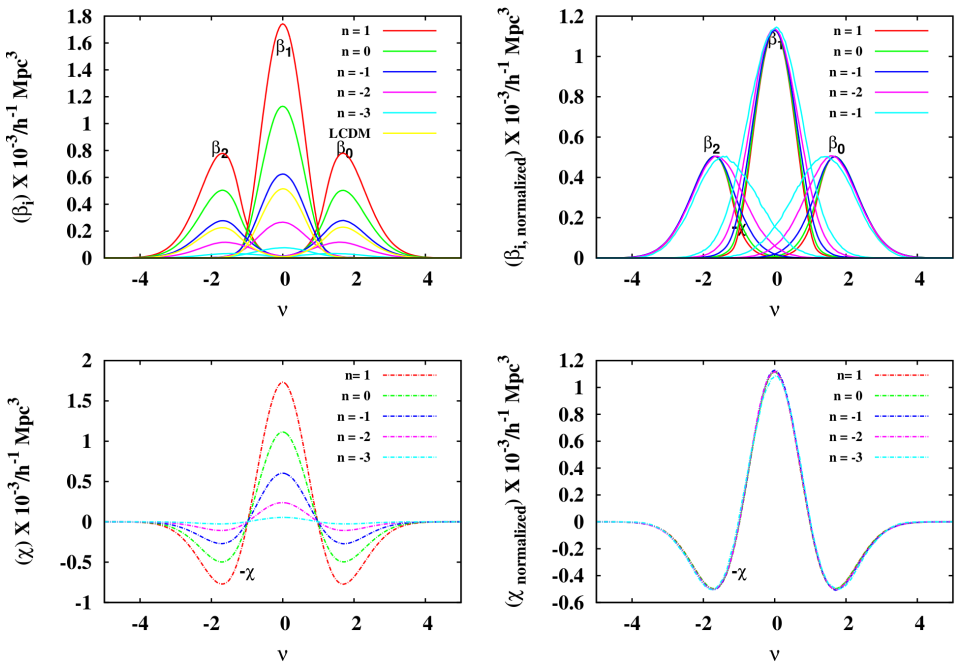


Figure 3.26 Betti numbers and Euler characteristic curves of Gaussian random fields. The curves are drawn for the LCDM model and for the power-law models as a function of the dimensionless density threshold ν. Top-left: Unscaled Betti numbers for the LCDM and power-law models. Top-right: Scaled Betti numbers for the power-law models. Bottom-left: unscaled Euler characteristic curves for the power law models. Bottom-right: Scaled Euler characteristic curves. The shape of the Betti number curves depend on the choice of the power spectrum, unlike the shape of the Euler characteristic curves.

The dependence of the Betti number curves on n is nearly exclusively confined to the range $|\nu| \leq \sqrt{3}$.

The bottom-left panel of Figure 3.26 shows the unscaled Euler characteristic curves for the power law models. The bottom-right panel presents the scaled Euler characteristic curves. The scaled curves fall on top of each other, indicating that the shape of the Euler characteristic curve is insensitive to the choice of power spectrum. This is unlike the Betti numbers, whose shapes show a characteristic dependence on the choice of the power spectrum. The dependence of Euler characteristic on the choice of the power spectrum is restricted to the expression for amplitude, through the variance term. Also note that the different Betti numbers dominate different regions of the Euler characteristic curve.

The extrema of the three 3D Betti numbers correspond to the three extrema of the Euler characteristic curve. Only for large thresholds of $|\nu| > 3$, β_0 and β_2 are almost equal to $-\chi$. This is because the absolute value of the Euler characteristic is very close to the number of excursion sets or peaks in the asymptotic limit of high density thresholds (Adler 1981; Bardeen et al. 1986). For thresholds as large as $\nu \sim 2$, there is

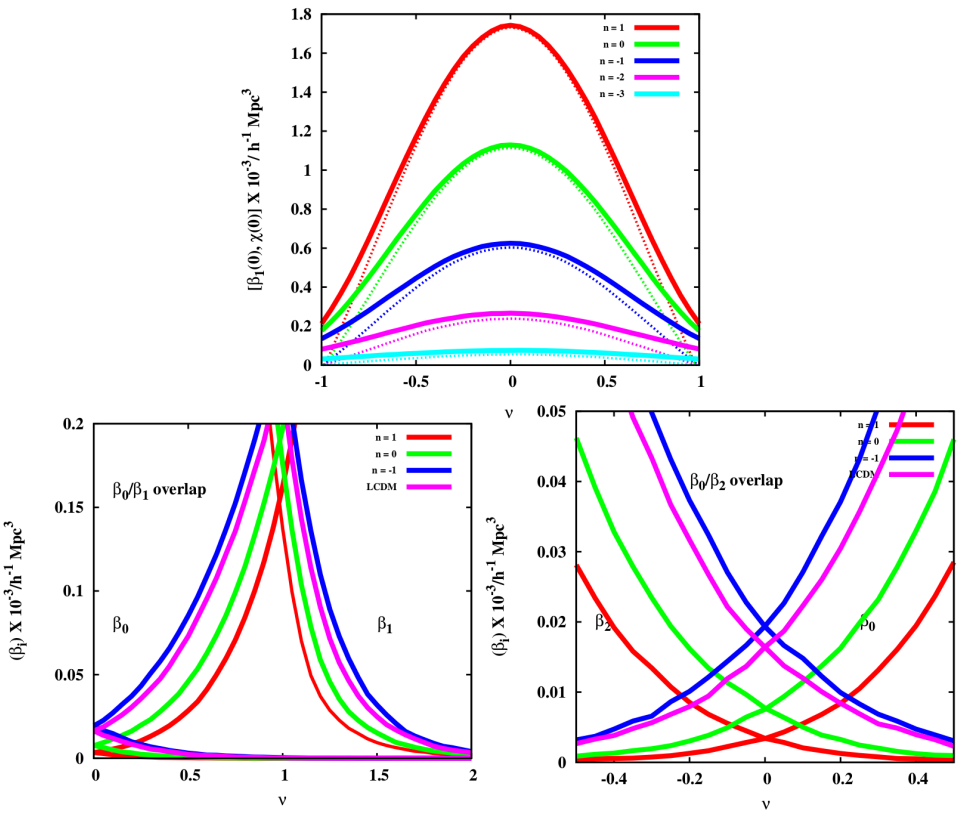


Figure 3.27 Zoom-in into the Betti numbers and Euler characteristic curves. Top: Zoom-in into the Betti number (solid) and Euler characteristic curves (dashed) around $\nu = 0$. The Euler characteristic curves have an amplitude slightly lower than the β_1 curves. Bottom-left: overlap between β_0 and β_1. Bottom-right: overlap between β_0 and β_2. In both the panels, it is evident that the overlap between the various Betti numbers spans across a substantial range of density threshold. The strength of the overlap between different topologies, measured by the relative number of different topological entities at a given threshold, increases for decreasing spectral index.

a significant contribution from β_1 to χ. For the $n = -3$ model, this is even as large as $10\% - 15\%$, as seen in the right panel of Figure 3.28. In the top panel of Figure 3.27 we see that the amplitude of χ is lower than the amplitude of β_1. The difference becoming larger as the spectral index decreases. It is an indication of the presence of a significant number of islands and voids at $\nu = 0$ for lower spectral indices.

The above observations can be related to the nature of the density fluctuation field as a function of spectral index. For higher spectral indices, there is significant power only at smaller scales. This results in high density peaks connected by low density saddles, giving the field a distinctly spiky appearance. These peaks get connected before they start forming tunnels and voids, resulting in a clear cut demarcation of meatball-, sponge- or cheese-like topology. As the spectral index

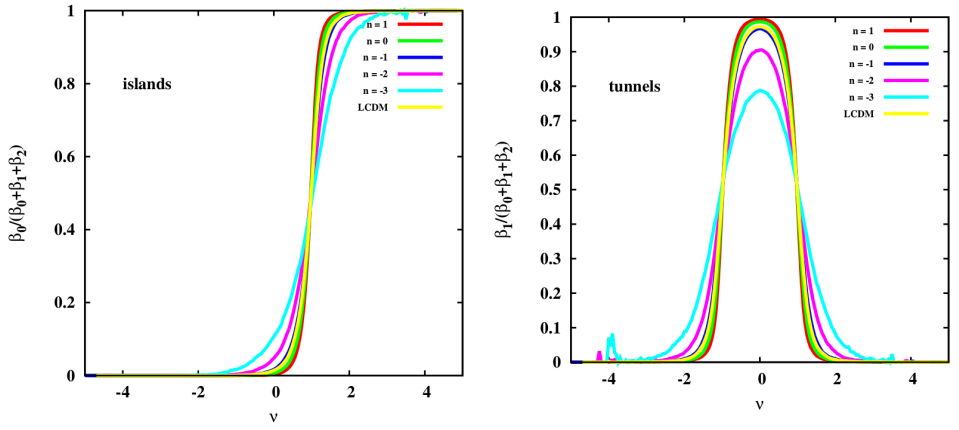


Figure 3.28 The fractional contribution of islands and tunnels to the sum of Betti numbers in all three dimensions for rms threshold range between $(-5\sigma : 5\sigma)$. Left: $\beta_0/(\beta_0+\beta_1+\beta_2)$ – fractional contribution of islands to the total sum of Betti numbers. Right: $\beta_1/(\beta_0+\beta_1+\beta_2)$ – fractional contribution of tunnels to the total sum of Betti numbers.

decreases, the demarcation diffuses. As the spectral index decreases, progressively more and more isolated islands contain additional topological holes of higher dimensions, at thresholds well before the manifold becomes a singly connected entity. This is reflected in the broadening and increased overlap of the Betti number curves, indicating an increase in the mixture of topology as the spectral index decreases. In contrast, the Euler characteristic curve does not have this dependence. As a result, this additional information about the inherent differences in the topological structure of the various power law models is not available from the Euler characteristic curves. It is clear from this discussion that the Betti numbers add extra information to the description of topology than that by the Euler characteristic.

The above remarks lead us to conclude the following. In general, only for positive spectral indices, it is feasible to describe the topology of the field as either *meatball-like*, or *sponge-like* or *cheese-like*. For negative spectral indices, the demarcation is not clear, except near the tails of the density distribution. The topology is an increasing mixture of the three types as the spectral index decreases.

3.6.2 Betti numbers: Scaling relation

The amplitudes of the unscaled Betti numbers depend on the value of the spectral index. The trend of the dependence of the maximum of the Betti number curves on the value of the spectral index is shown in Figure 3.29. The amplitudes are fit to an exponential of the form $f(n) = A_0 e^{n/\tau}$, with the decay parameter $\tau = 2$. This suggests that the amplitude of all the Betti number curves decrease roughly exponentially as the value of spectral index decreases.

The amplitudes for β_0, β_1 and β_2 scale the same as a function of the smoothing length R_g. We fit the maximum of the peaks to the function $\beta_i^{max} = A_0 R_g^{\tau}$. Table 3.2

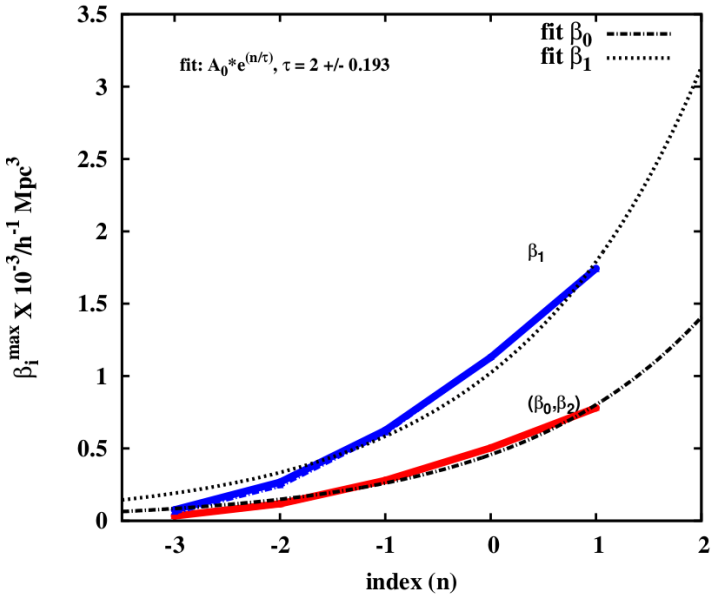


Figure 3.29 Amplitude of Betti numbers as a function of the spectral index. The amplitudes are fit to an exponential function of the form $a \cdot \exp \nu$. The χ^2 of the fits is 0.007 and 0.033 for the islands/voids and tunnels respectively.

presents the values of the parameters of fit for the various models. The values of τ suggest that the Betti numbers scale as $R_g^{(n+3)}$, which is also proportional to σ_0^2. This is the same as the scaling law for genus. The reduced χ^2 of the fit is 0.003. Park et al. (2013) show that the scaling relation of the Betti number curves is very close to that of the genus curve, which is $gR_G^3 \propto (n+3)^{3/2}$ (Hamilton et al. 1986), but not exactly the same.

3.6.3 Peaks vs. Islands: the Gaussian case

In Section 3.3.5, we noted that there is a difference between peaks and islands. The former are the local maxima, while the latter grow depending on the density thresh-

index (n)	$\tau_{(\beta_0,\beta_2)}$	$\tau_{(\beta_1)}$
1	4.027 ± 0.01	3.99 ± 0.01
0	3.012 ± 0.018	2.99 ± 0.02
-1	2.001 ± 0.01	1.98 ± 0.015
-2	1.015 ± 0.05	0.99 ± 0.01
-3	0.0015 ± 0.002	0.001 ± 0.002

Table 3.2 Scaling of Betti numbers as a function of smoothing radius for the power-law models with spectral indices $n = 1, 0, -1, -2$ and -3. The maxima of Betti numbers is fit to the function $\beta_i^{max}(R_g) = A_0 R_g^{\tau}$. The values of τ suggest that the Betti numbers scale as $R_g^{(n+3)}$, which is the same as the scaling law for genus. Reduced χ^2 of the fit is 0.003.

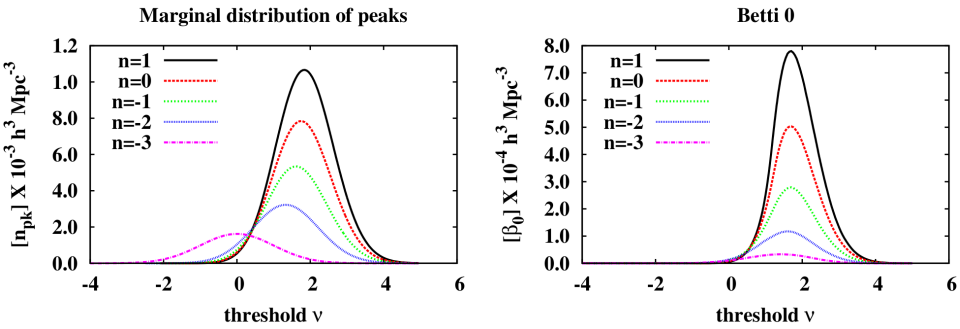


Figure 3.30 Marginal distribution of peaks, and the Betti numbers. The left and the right columns plot the marginal distribution of peaks and the zeroth Betti number respectively. The zeroth Betti number counts the number of isolated islands at a particular density threshold. The curves for both the quantities show a characteristic dependence on the index of the power spectrum. Note also that the location of peaks shifts towards lower density thresholds for both the quantities. This effect is strong for the marginal distribution of peaks, but marginal for the Betti numbers.

old ν. An island may contain multiple peaks. In fact, as an island grows it might get arbitrarily complicated, acquiring tunnels and even voids, while always staying connected. But in the asymptotic limit of high ν, every island will contain only one peak.

Bardeen et al. (1986) derive the marginal number distribution of peaks for Gaussian random fields, as a function of the dimensionless density threshold ν

$$\mathbb{N}_{pk}(\nu)\, d\nu = \frac{1}{(2\pi)^2 R_\star^3} e^{-\nu^2} G(\gamma, \gamma\nu). \tag{3.33}$$

Here, the function $G(\gamma, \gamma\nu)$ is a fitting function (see Appendix C.1.1 for the exact form). For the power law models, the parameters γ and $R_\star$ are related to the various moments of the power spectrum, the value of the spectral index n, and the co-moving filtering radius R_f.

Figure 3.30 plots the marginal distribution of peaks and the Betti numbers for the 3D Gaussian fields. The left and the right panels plot the marginal distribution of peaks and the zeroth Betti number β_0 respectively. The curves are drawn as a function of the dimensionless density threshold ν. The marginal distribution of peaks counts the number density of peaks in the neighborhood $(\nu, \nu + d\nu)$. The zeroth Betti number counts the number of isolated islands at ν. The curves for both the quantities show a characteristic dependence on the index of the power spectrum. The curves of both the quantities also show different characteristics from each other. This indicates that they measure different features associated with the topology of the density distribution. The number density of peaks differs from the zeroth Betti number because the zeroth Betti number also depends on the number distribution of 2-saddles. From the graphs, it is evident that the location of peaks shifts towards lower density thresholds for both the quantities, as the spectral index decreases. This effect is

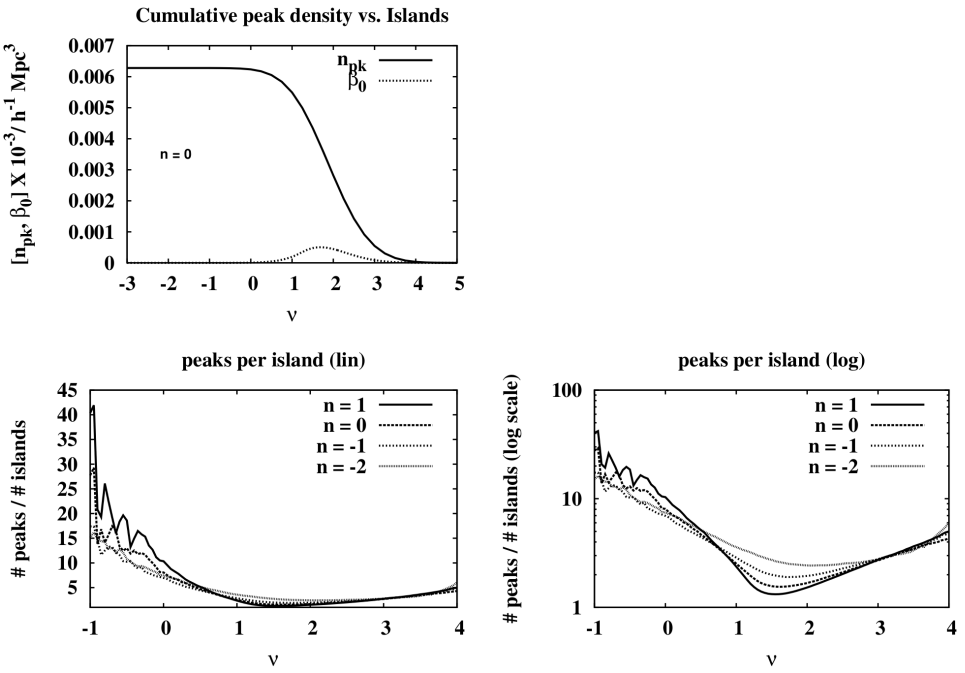


Figure 3.31 Top: Cumulative number distribution of islands and peaks as a function density threshold for the $n = 0$ model. Bottom-left: Number of peaks per island as a function of the density threshold, for the different power law models. Bottom-right: The number of peaks per island presented in the logarithmic scale.

strong for the marginal distribution of peaks, but small for β_0. Note also the orders of magnitudes of separation between the amplitudes of both the quantities.

For the $n = 1$ model, the small scales are dominant. In terms of the structures in the density fluctuation field, this means that the number of small scale peaks of high amplitude is large. They are also separated by low-density saddles. There is no discernible large scale feature in the density field. As the index of the power spectrum decreases, the power shifts to large scales. The small scale peaks are separated by saddles occurring at relatively high density thresholds. As a result the features formed by these peaks and saddles are of low persistence. It is also accompanied by a decrease in the amplitude of the global maximum of the field. As noted earlier, this is because the variance of the density field in the box decreases with decreasing spectral index: $\sigma_0 \propto R^{-(n+3)}$. This phenomenon is reflected in the curves of the marginal distribution of peaks in the Figure 3.30. For a larger n, the number distribution of peaks attains its maximum at a larger density threshold compared to a smaller n. For the $n = 1$ model, the maxima is located as high as $\nu = 2$, and the number density rapidly starts approaching zero near $\nu = 0$. In contrast, for the $n = -3$ model, the maximum is located at $\nu \sim 0$, and there are significant number of peaks even below $\nu = 0$. As noted earlier, this is a direct reflection of the fact that there are progres-

sively more number of peaks for lower thresholds, as the spectral index decreases (Bardeen et al. 1986). In contrast, the location of the maxima of β_0 curves shows a marginal dependence on the value of spectral index.

For very high values of rms density threshold, as long as the peaks do not start merging, we expect the cumulative number density of peaks to be equal to the number density of islands. This is confirmed in the left panel of Figure 3.31 where we present the cumulative number density of peaks n_{pk} and the number density of islands β_0 per unit volume, as a function of ν for the $n = 0$ model. The cumulative number density of peaks equals the number density of islands asymptotically for very large rms density thresholds. The equivalence starts breaking down rapidly at thresholds even as high as $\nu \sim 4$. This is attributed to the fact that for high thresholds all the peaks represent disconnected regions almost surely (Bardeen et al. 1986), while they start connecting up and forming complex topology as the threshold decreases.

In the middle and the right panel of Figure 3.31, we present the number of peaks per island as a function of the dimensionless density threshold ν. The vertical axis of the middle panel is plotted on a linear scale, while it is in a logarithmic scale for the right panel. The number of peaks per island reaches a minimum within $1 \leq \nu \leq 2$ for all the models. The location of the minimum shifts to lower density thresholds for decreasing spectral index. The value at the minimum increases for decreasing spectral index, indicating that the number of peaks per island around the minimum decreases with increasing spectral index. For lower density thresholds, the trend is reversed. There are more number of peaks per island for increasing spectral index. In this connection, it is important to note that the number of peaks per island depends on two quantities : the number distribution of peaks themselves, as well as the number distribution of the 2-saddles, as a function of the density threshold. Both these quantities behave differently in general (Bardeen et al. 1986; Pogosyan et al. 2009). The situation is further complicated by the fact that not all 2-saddles at a given density threshold join two disconnected peaks. A fraction of them is also responsible for connecting two or more already connected peaks, thereby forming loops or tunnels (Edelsbrunner & Harer 2010; Feldbrugge et al. 2015). Further, the fraction of 2-saddles that join two isolated objects would be, in general, a function of the density threshold (also see Feldbrugge et al. (2015) for a semi-analytic framework describing this).

3.7 Minkowski functionals of Gaussian random fields

In principle, the Minkowski functionals provide complementary geometric information to the topological analysis described in this chapter. As such, it is of importance to assess the correspondence between these quantities.

There are $(d+1)$ Minkowski functionals, $Q_k((k = 0, \ldots, d)$, for a d-dimensional manifold (Mecke et al. 1994; Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). Predominantly geometric in nature, the first four Minkowski functionals ($d \leq 3$) are the volume, surface area, integrated mean curvature or total contour length, and the Gaussian curvature. Within the context of this chapter,

we assume a 3-dimensional space **x**. We calculate the Minkowski functionals for the manifold M_ν, which is the subset of the region **x** with density above the threshold ν.

The volume functional $Q_0(\nu)$ is the fractional volume of the regions with density above the threshold ν, normalized by the total volume of the region V. It is given by the volume integral of the Heaviside step function Θ, normalized by the whole volume V (Schmalzing & Buchert 1997)

$$Q_0(\nu) = \frac{1}{V}\int_V \Theta(\nu - \nu(\mathbf{x})). \tag{3.34}$$

The other Minkowski functionals of M_ν can be calculated by appropriate surface integrals of the boundary of M_ν, denoted by $\partial(M_\nu)$. The second Minkowski functional, or the area functional, is given by

$$Q_1(\nu) = \frac{1}{6V}\int_{\partial M_\nu} d^2S(\mathbf{x}). \tag{3.35}$$

The third and the fourth Minkowski functionals, namely the integrated mean curvature functional or the total contour length, and the Gaussian curvature, involve the inverse of the radii of curvatures R_1 and R_2 of the surfaces oriented towards lower density values. The integrated mean curvature functional is given by

$$Q_2(\nu) = \frac{1}{6\pi V}\int_{\partial M_\nu} d^2S(\mathbf{x})[\kappa_1 + \kappa_2], \tag{3.36}$$

and the Gaussian curvature is given by

$$Q_3(\nu) = \frac{1}{4\pi V}\int_{\partial M_\nu} d^2S(\mathbf{x})[\kappa_1\,\kappa_2], \tag{3.37}$$

where $\kappa_1 = 1/R_1$ and $\kappa_2 = 1/R_2$.

3.7.1 Analytical expressions for Minkowski functionals of Gaussian random fields

For Gaussian random fields, the expected value of the first four Minkowski functionals of the excursion sets have known analytical expressions (Tomita 1993; Schmalzing & Buchert 1997)

$$
\begin{aligned}
Q_0(\nu) &= \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{1}{\sqrt{2}}\nu\right), \\
Q_1(\nu) &= \frac{2}{3}\frac{\lambda}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\nu^2\right), \\
Q_2(\nu) &= \frac{2}{3}\frac{\lambda^2}{\sqrt{2\pi}}\nu\exp\left(-\frac{1}{2}\nu^2\right), \\
Q_3(\nu) &= \frac{\lambda^3}{\sqrt{2\pi}}(\nu^2-1)\exp\left(-\frac{1}{2}\nu^2\right).
\end{aligned} \tag{3.38}
$$

where $\lambda = \sqrt{|\xi''(0)|/[2\pi\xi(0)]}$, and the function $\Phi(x) = \int_0^x \mathrm{d}te^{(-t^2)}$ is the standard error function. This makes them an ideal tool for model discrimination.

3.7.2 Computing the Minkowski functionals

Suppose the density at a location $\mathbf{x}$ is specified as $u(\mathbf{x})$. The volume functional is computed directly from Equation 3.34. Replacing the surface integration in Equations 3.35, 3.36 and 3.37 with a spatial mean over the whole volume, we get (Koenderink 1984)

$$
\begin{aligned}
Q_1(\nu) &= \frac{1}{6V}\int_V d^3x\,\delta(\nu\sigma - u(\mathbf{x}))|\nabla u(\mathbf{x})|, \\
Q_2(\nu) &= \frac{1}{6\pi V}\int_V d^3x\,\delta(\nu\sigma - u(\mathbf{x}))|\nabla u(\mathbf{x})|[\kappa_1+\kappa_2], \\
Q_3(\nu) &= \frac{1}{6\pi V}\int_V d^3x\,\delta(\nu\sigma - u(\mathbf{x}))|\nabla u(\mathbf{x})|[\kappa_1\,\kappa_2],
\end{aligned} \tag{3.39}
$$

where δ is the Dirac delta function. The local radii of curvature can be expressed in terms of geometric invariants known as the *Koenderink invariants*. They are formed from the first and the second derivatives of the density field $u(\mathbf{x})$ (Koenderink 1984; Kerscher et al. 1996)

$$
\kappa_1+\kappa_2 = \frac{\epsilon_{ijm}\epsilon_{klm}\,u_{,i}u_{,jk}u_{,l}}{(u_{,n}u_{,n})^{3/2}};\ \kappa_1\kappa_2 = \frac{\epsilon_{ijk}\epsilon_{lmn}\,u_{,i}u_{,l}u_{,jm}u_{,kn}}{2(u_{,p}u_{,p})^2}, \tag{3.40}
$$

where $u_{,i}$ and $u_{,ij}$ are the first and second spatial derivatives in the i and j direction, and ϵ_{ijk} is the Levi-Civita tensor. Equation 3.39 can be performed computationally, when the inverse of radii of curvature is replaced by the form given in Equation 3.40. Kerscher et al. (1996) and Schmalzing & Buchert (1997) develop a code to compute the Minkowski functionals of density fields specified on a regular grid, based on the above prescription [3]. We adapt the code to perform the computation of Minkowski

[3]The original code is available at `http://www.physik.uni-muenchen.de/sektion/lswagner/buchert_software.html`

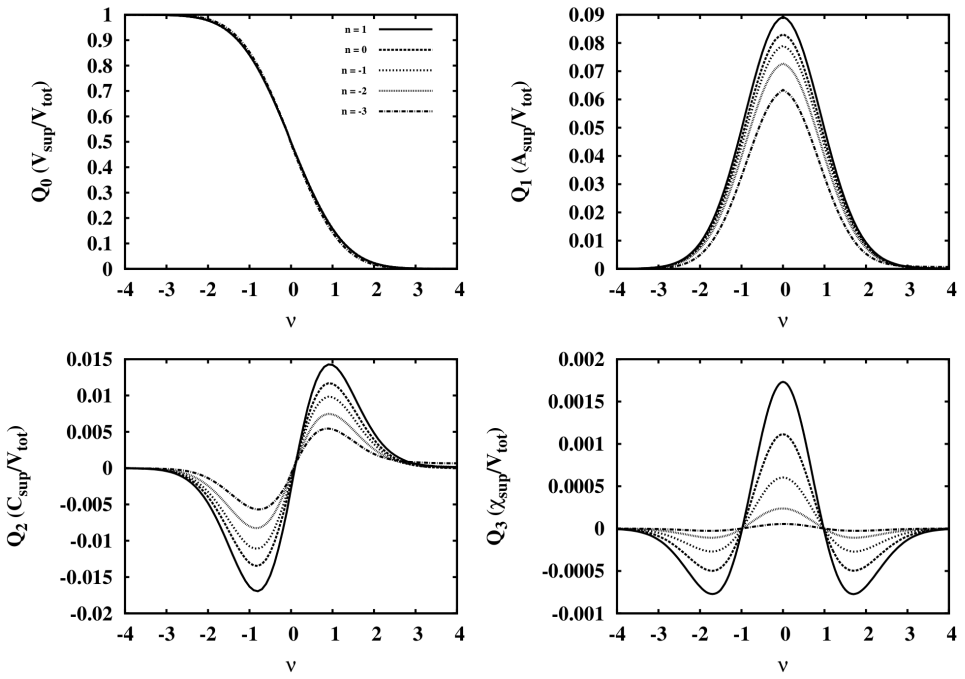


Figure 3.32 Minkowski functionals of the LCDM and the power-laww models. Top left: Volume functional. Top right: Area functional. Bottom left: Integrated mean curvature functional. Bottom right: Euler characteristic or Gaussian curvature functional. All the functionals are normalized by the total volume of the simulation box. The volume functional is invariant with respect to the choice of the power spectrum. The amplitude of the area, contour length and Euler characteristic shows a dependence on the choice of the power spectrum.

functionals of the models described in this chapter.

3.7.3 Minkowski functionals of the models

Figure 3.32 presents the graph of the Minkowski functionals for the power-law models. The graphs are averaged over 100 realizations. The quantities are plotted as a function of the density threshold ν. The top-left panel presents the volume functional Q_0,the top-right panel plots the area functional Q_1, the bottom-left panel plots the integrated mean curvature functional Q_2, and the bottom-right panel plots the Euler characteristic Q_3. All the functionals are normalized by the total volume of the simulation box. The fractional volume Q_0 is invariant with respect to the choice of the power spectrum. All the other functionals show a systematic dependence on the choice of the power spectrum. The amplitude of the graphs of the area functional, the integrated mean curvature functional, and the Euler characteristic decreases monotonically with the decrease in the index of the power spectrum.

Figure 3.33 presents the rescaled Minkowski functional curves. In this figure, the graphs for the rest of the power-law models have been scaled to the amplitude of

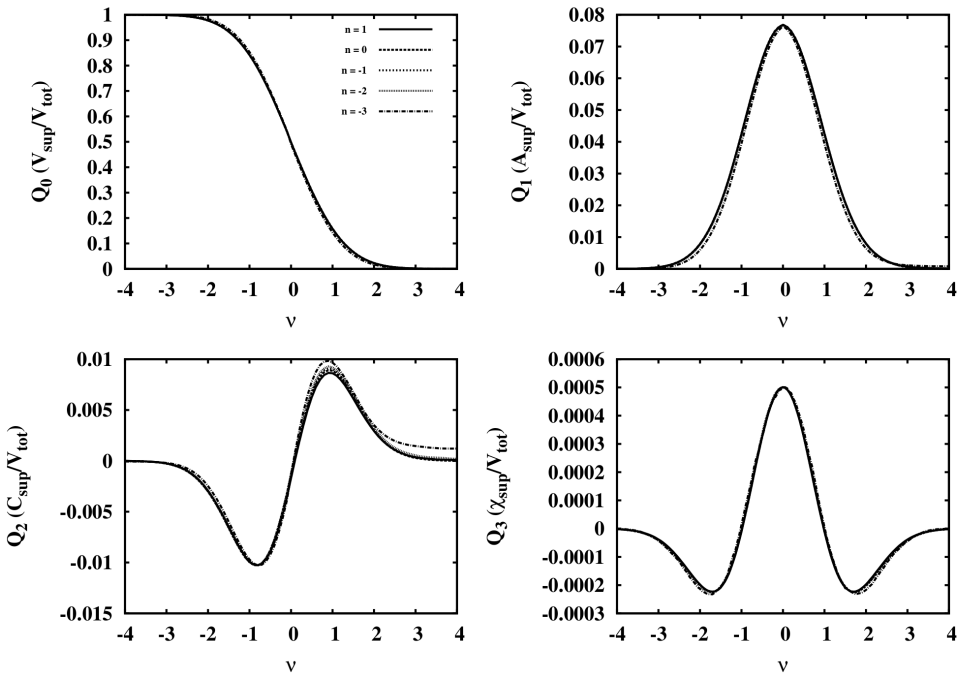


Figure 3.33 Scaled Minkowski functionals of the LCDM and the power-law models. Top left: Volume functional. The power-law curves are scaled to the amplitude of the LCDM curve. Top right: Area functional. Bottom left: Integrated mean curvature functional. Bottom right: Euler characteristic. All the functionals are normalized by the total volume of the simulation box. The shape of the curves is invariant with respect to the choice of the power spectrum.

the curve of the $n = 0$ model. The shape of the rescaled graphs falls neatly on top of each other. This indicates that the shape of the Minkowski functional curves is independent of the choice of the power spectrum. This observation is in line with the Equation 3.38. The dependence on the choice of the power spectrum comes in only through the amplitude term. This dependence is parametrized in terms of λ, which is a function of the correlation function or equivalently the power spectrum.

That the shape of the Minkowski functional curves is independent of the choice of the power spectrum is an important observation, when seen in comparison to the shape of the Betti number curves, which show a characteristic dependence on the choice of the power spectrum. We present a detailed analysis of the Betti numbers with respect to the Euler characteristic in Section 3.6. This indicates that the Betti numbers are potentially more discriminatory than the Minkowski functionals.

Recall that the ratio of the Minkowski functionals are a crude indicator of the morphological properties of manifold (Sahni et al. 1998). For example, a high surface area to volume ratio indicates a more pancake like morphology of structures. The reverse indicates a more filamentary morphology. Figure 3.34 presents the ratios of the first three Minkowski functionals with respect to each other. In the left panel, we

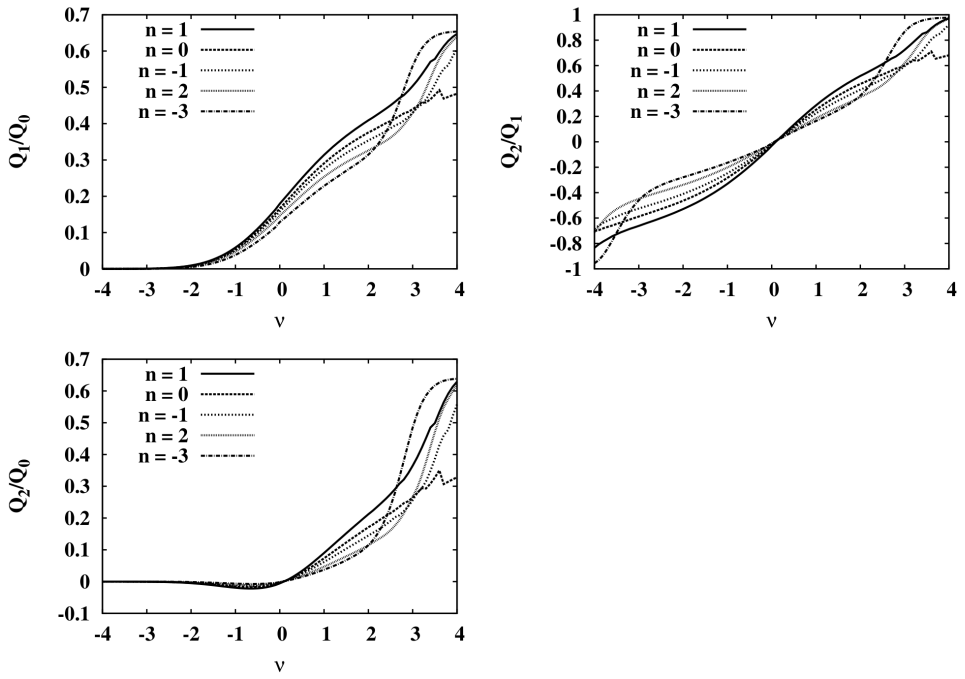


Figure 3.34 Ratios of Minkowski functionals. Left: The ratio Q_1/Q_0, denoting the ratio between the total occupied volume and the total surface area corresponding to the occupied volume. Middle: The ratio Q_2/Q_1, denoting the ratio between the total length of contours and the total surface area. Right: The ratio Q_2/Q_0, denoting the total length of contours per unit volume occupied. The curves are drawn with respect to the dimensionless density threshold ν.

present the ratio Q_1/Q_0, denoting the ratio between the total surface area and the total occupied volume. In the middle panel, we present the ratio Q_2/Q_1, denoting the ratio between the total length of contours and the total surface area. In the right panel, we present the ratio Q_2/Q_0, denoting the total length of contours per unit volume occupied. The curves are drawn with respect to the dimensionless density threshold ν.

All the three quantities increase monotonically for increasing density thresholds. However, they also show a characteristic dependence on the choice of the power spectrum. Note that the curves for the negative power spectra increase more steeply towards the extremes of the density threshold, and flatten out as the threshold moves to further extremes. This is in contrast to the behavior for the $n = 1$ and the $n = 0$ models, for which the curves show no such behavior. For all the models, the surface area to volume ratio is high for high density thresholds. It indicates that the structures are more flattened for high thresholds. Interesting is the sharp rise in the value for the negative spectra. This indicates that at very high thresholds, the structures in the $n = -3$ model are the most flat. This ties in with the observation that for the $n = -3$ model, there are large scale structures that have low amplitude. This gives

rise to the overall flattened characteristics of the density field. The large structures are a consequence of significant powers at those scales.

In summary, the Minkowski functionals characterize the geometric properties of the manifold predominantly. The connection to topology comes through the Euler characteristic. Hence, the Minkowski functionals maybe seen as complimentary to the topological descriptors such as persistence and Betti numbers. The Minkowski functionals, together with the information on the homology and persistence of a manifold, provide a richer and more comprehensive morphological and topological information about the manifold.

3.7.4 Betti numbers vs Minkowski functionals

As we learnt in the previous sections, the Betti numbers are topological quantities. They measure topology by assessing the number of independent holes in the different dimensions. On the other hand, the Minkowski functionals are primarily morphological measures, the exception being the Minkowski functional Q_3, or the Euler characteristic, χ. The first three Minkowski functionals are associated with the volume (Q_0), surface area (Q_1) and the integrated mean curvature length (Q_2) of the manifold. However, an important question one may ask is if the Betti numbers and the Minkowski functionals convey different information about the manifold characteristics. With a view to investigate this, we assess the correspondance between the Betti numbers and the Minkowski functionals.

Figure 3.35 presents the Betti numbers plotted against the various Minkowski functionals. The top-left panel of the figure plots β_0 on the vertical axis against Q_0 on the horizontal axis, and so on. We notice that almost all the pairs of quantities exhibit a degeneracy. For example, in the top-left panel, we notice that there are two values of Q_0 for which the value of β_0 is the same. The exception is the peak of the curve, at which β_0 is associated with a unique value of Q_0. The only exception to this trend of degeneracy is the middle panel of the middle row, where we plot β_1 against Q_1. The curve is monotonic, indicating that β_1 and Q_1 behave in a similar fashion. In general, a monotonic curve between any two plotted quantities indicates a similar behaviour of the quantities.

Another simple method to investigate whether the Betti numbers and Minkowski functionals present similar information about the manifold characteristics is to plot the ratio of the Betti numbers to the Minkowski functionals. In certain cases, the ratio of a given Betti number to a given Minkowski functional can be associated with particular features of the manifold. For example, the ratio β_0/Q_0 represents the number of isolated objects per unit occupied volume. Note that the occupied volume is different from the total volume of the manifold. The total volume is a constant, while the occupied volume is a function of the density threshold. Similarly, the ratio β_1/Q_1 indicates the number of independent tunnels per unit surface area. This may be regarded equivalent to the information on the genus of the manifold.

Figure 3.36 presents the ratio of the various Betti numbers to the various Minkowski functionals, as a function of the density threshold ν. The plots are presented for the different power law models. We notice a dependence of the quantities on the choice of the spectral index. It is important to note that a constant, or a monotonically in-

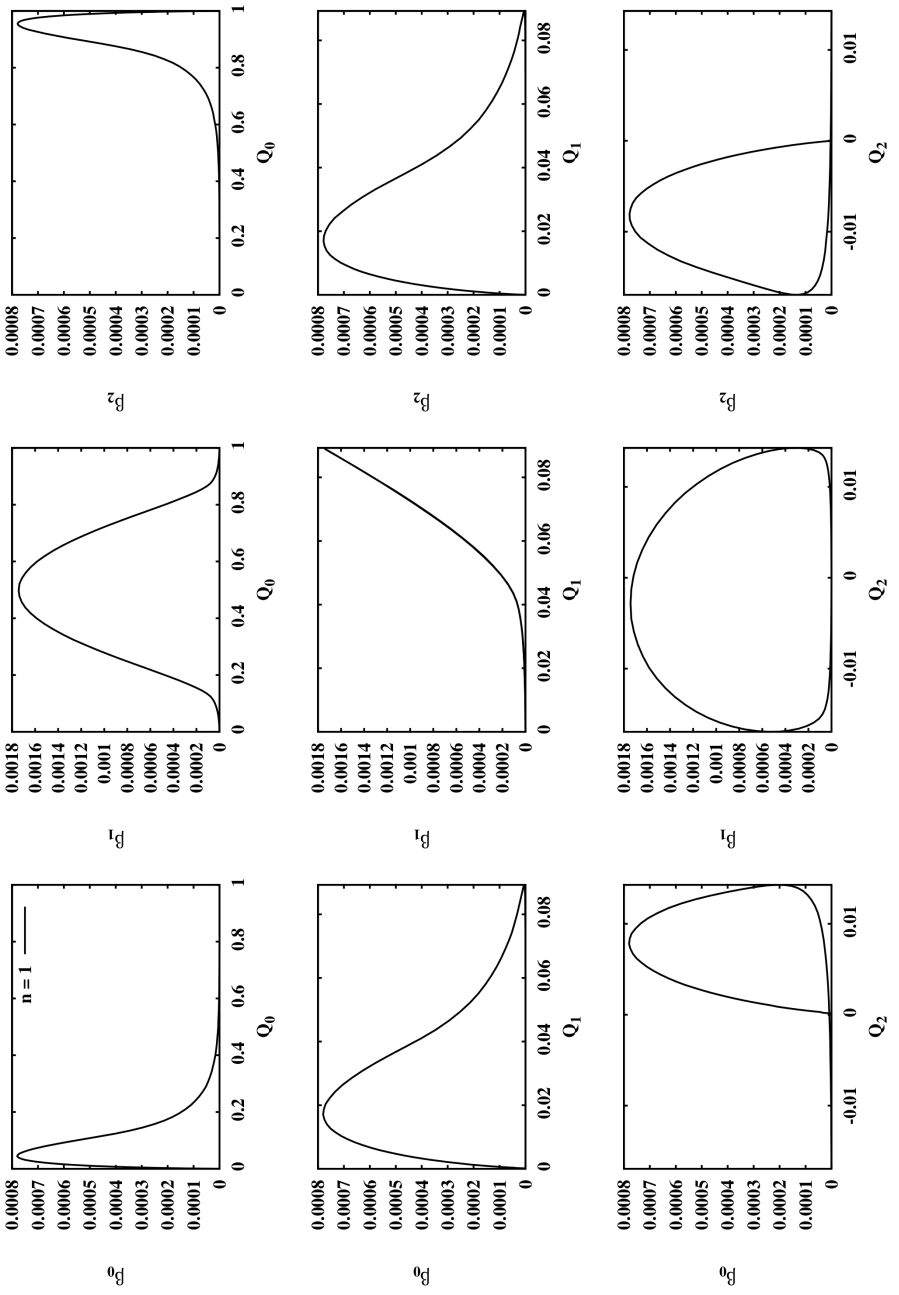


Figure 3.35 Betti numbers vs. the Minkowski functionals.

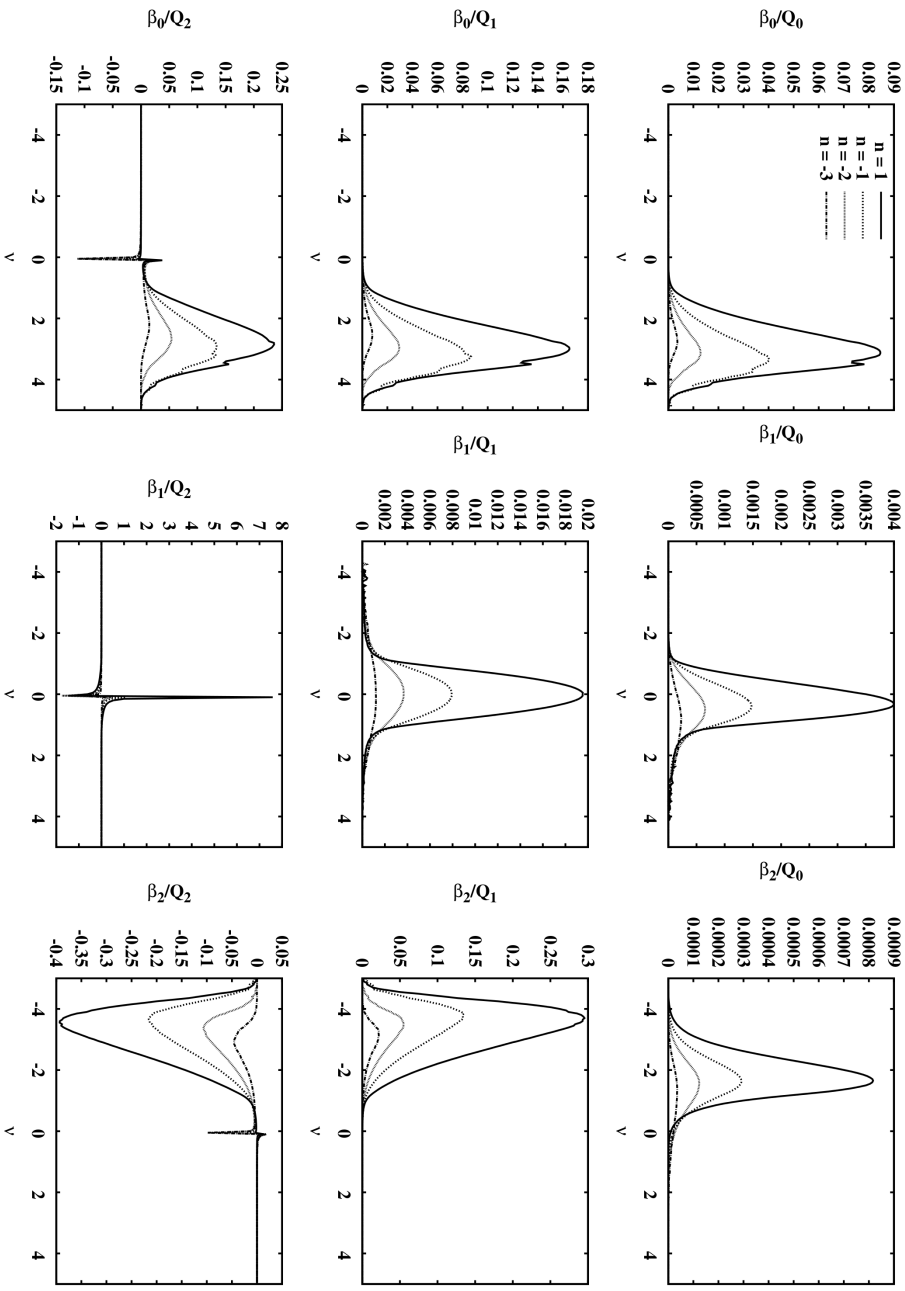


Figure 3.36 Ratio of Betti numbers to the various Minkowski functionals. The graphs as plotted for a range of density threshold.

creasing or decreasing curve indicates quantities with similar characteristics. This is because in the case of a constant or a monotonic ratio between two quantities, one differes from the other by only a factor. We notice that none of the pair of quantities exhibit a monotonic ratio. This indicates crudely that the Betti numbers and the Minowski functionals behave differently from each other in general.

3.8 Discussions and conclusions.

This chapter analyzes the topology of Gaussian random fields in a hierarchical fashion through concepts emanating from homology, persistence and Morse theory. Such a topological analysis of Gaussian random fields is interesting in the cosmological context given that the density perturbations in the early Universe are predicted to be highly Gaussian in nature.

A hierarchical topological characterization of the cosmic mass distribution has a rich potential of unlocking tools for studying the nature of the cosmic mass distribution, especially given the fact that structure formation in the Universe proceeds in a hierarchical fashion.

We study the power-law and the LCDM model using using a variety of topological measures. We demonstrate that a combination of intensity, difference and ratio maps express the topological information content in greater detail than the traditional descriptors like the Minkowski functionals.

First, we delve into the topological description of 1D Gaussian functions. This is done in order to relate the visual features of the density distribution to the topological description, and assess how the topological description reflects the features of the density distribution. We establish that the features of the persistence diagram are a direct reflection of the singularity structure of the function.

Having assessed the topology of the 1D functions, we then delve into an analysis of the 3D fields. We do this for the power law models and the LCDM model. First we present the intensity, difference and the ratio maps of the power law models, as well as the LCDM models. We demonstrate that the difference and ratio maps present the differences between the topology of two models in comparison at a more detailed level. The indication is that the intensity, difference and ratio maps are highly sensitive to the parameters of the model, and may therefore be used to discriminate between various models.

We quantify the intensity maps by a decomposing them into marginal and cumulative distributions of the mean-density and persistence of the topological holes in different dimensions. In this context, we establish that the distribution functions show a characteristic dependence on the index of the power spectrum. We also show that the marginal distribution of mean density of the holes follow a Gaussian distribution approximately. The marginal distribution as a function of persistence indicates a Gamma distribution.

Thereafter, we proceed to compute the Betti numbers for the models, and compare them with the Euler characteristic of the models. Recalling that the Euler characteristic is the alternating sum of Betti numbers, we also demonstrate that the Betti numbers of Gaussian random fields contain strictly extra information compared to

the Euler characteristic. In particular, we show that the shape of the Betti number curves is dependent on the choice of the index of the power spectrum. In contrast, the curves of shape of the Euler characteristic curve is insensitive to the choice of the power spectrum.

A crucial observation is that the topology of the manifold is not strictly either meatball-like, sponge-like or cheese-like for the density fields. This is demonstrated in details in the Section 3.6, where we find that there are substantial regions of overlap between the various kinds of topologies enumerated above. This claim is reinforced by the scaled Betti number curves that show a substantial overlap across a range of density thresholds. The overlap is the strongest for a lower spectral index, and decreases monotonically with increasing spectral index. We note that such information may not be available through the Euler characteristic curves.

Further, we devote a section to understanding the relationship between peaks and islands. We study their relationship by computing the marginal distribution of peaks and the zeroth Betti number. The quantities behave differently, except for asymptotically high density thresholds. This is established by plotting the cumulative number distribution of peaks with respect to β_0. In this process, we also compute the number of peaks per island as a function of the density threshold.

Finally, we present an analysis in terms of the Minkowski functionals of the models. The Minkowski functionals measure the morphological properties of the manifold, hough there are connections to topology through the $(d+1)$-th functional. Most importantly, we show that the shape of the Minkowski functionals is independent of the choice of the power spectrum, unlike the Betti numbers. We also show that the ratio of the Minkowski functionals, which denote the shape characteristics of the density field, show a dependence on the choice of the power spectrum. The indication is that the shapes become progressively more flattened as the spectral index decreases. We conclude that the Minkowski functionals are complimentary to topological measures like persistence and the Betti numbers. A detailed information about all of them presents a comprehensive picture of the morphological as well as the topological attributes of the manifold.

Given the extra information content of persistence homology and homology, as quantized by the Betti numbers, they will prove to be key concepts for a more profound understanding of the topology in the future. As an example, it would be interesting to analyze the persistence and homology of the temperature fluctuations of the Cosmic microwave background, as recently measured in extraordinary details by PLANCK satellite.

4

Felix: A Topology based Framework for Visual Exploration of Cosmic Filaments

— Pratyush Pranav—

I know simply that the sky will last longer than I. – *Albert Camus, The Myth of Sisyphus and Other Essays*

4.1 Introduction

At scales from a megaparsec to a few hundred megaparsecs[1], the universe has a web-like appearance. In the *cosmic web* (Bond et al. 1996; van de Weygaert & Bond 2008), galaxies, intergalactic gas, and dark matter have aggregated in an intricate wispy spatial pattern marked by dense compact *clusters*, elongated *filaments* and sheetlike *walls*, and large near-empty void regions. The filaments, stretching out as giant tentacles from the dense cluster nodes, serve as transport channels along which mass flows towards the clusters. They surround the flattened walls, which are tenuous, membrane-like features in the cosmic mass distribution.

All structures and objects in the universe emerged out of primordial fluctuations that were generated during the *inflationary era*, moments after its birth, as the universe underwent a rapid phase of expansion (Guth 1981; Linde 1982a,b). The quantum fluctuations generated during this phase manifest themselves as fluctuations in the temperature of the cosmic microwave background (Spergel et al. 2007; Komatsu et al. 2011; Planck Collaboration et al. 2015). The gravitational growth of these density and velocity perturbations has resulted in the wealth of structure that we see in the Universe. The web-like patterns mark the transition phase from the primordial Gaussian random field to highly nonlinear structures that have fully collapsed into halos and galaxies. As our insight into the complex structural pattern of the cosmic web has increased rapidly over the past years, it has become clear that the cosmic-web contains a wealth of information on a range of cosmological and astronomical aspects and processes.

An important illustration of the cosmological significance of the cosmic web concerns its dependence on the nature of dark energy and matter, the dominant but as yet unidentified forms of energy and matter in the Universe. One telling example of this is the recent realization that cosmic voids are sensitive and useful probes of the nature of dark energy and dark matter and testing grounds for modified gravity theories (Ryden & Melott 1996; Schmidt et al. 2001; Park & Lee 2007a; Lavaux & Wandelt 2010; Bos et al. 2012; Lavaux & Wandelt 2012; Clampitt et al. 2013; Sutter et al. 2014b). As the cosmic web is first and foremost defined and shaped by the gravitationally dominant dark matter, it would be of considerable importance to be able to obtain detailed maps of dark matter distribution. In recent years, great strides have been made towards this goal as gravitational lensing of distant galaxies and objects by the dark matter have enabled an increasingly accurate view of its spatial distribution (Tyson 2000; Kneib & Natarajan 2011). Initial efforts concentrated on the detection and mapping of the deep potential wells of the nodes in the cosmic web, i.e., of galaxy clusters. Recent results have opened the path towards the mapping of filaments via their lensing effect on background sources (Dietrich et al. 2012). The identification of the structural components of the cosmic web is also important for our understanding of the relation between the formation, evolution, and properties of galaxies and the structural environment of the cosmic web. A direct manifestation of this is the generation of the angular momentum of galaxies. This is a product of

[1] A parsec is the standard unit of measurement of distances in the cosmos. A parsec is 3.26 times the *light-year*, the distance light covers in a year. A megaparsec is a million parsecs, the typical scale of measurement of size of the large scale structures in the universe.

the torqueing by the large-scale tidal force field (Hoyle 1951b; Peebles 1969; White 1984). While these force fields are also the agents for the formation and shaping of filaments, we would expect that this results in the alignment of the spin axis of galaxies with respect to cosmic filaments (Lee & Pen 2000; Jones et al. 2010; Tempel et al. 2013).

The identification, description, and characterization of the elements of the cosmic web is a non-trivial problem. Several characteristics of the mass distribution in the cosmic web have made it an extremely challenging task to devise an appropriate recipe for identifying them:

- The cosmic web is a complex spatial pattern of connected structures displaying a rich geometry with multiple morphologies and shapes.
- There are no well-defined structural objects at a single spatial scale or within a specific density range. Instead, elements of cosmic web are found at a wide range of densities and spatial scales. This is a consequence of the hierarchical evolution of structure formation in the universe, such that smaller high-density structures merge to form larger objects.
- There is a clear anisotropy in the structures of the cosmic web, a consequence of gravitational instability. The structures in the cosmic web exhibit elongated and flattened characteristics.

The attempts to analyze the structure of the cosmic web have a long history. The absence of an objective and quantitatively accurate procedure for identifying and isolating the components of the cosmic web has been a major obstacle in describing it. In recent years, more elaborate and advanced techniques have been developed to analyze and describe the structural patterns in the cosmic web. Nonetheless, a consensus on the proper definition of filaments is yet to be achieved. In the subsequent subsection 4.1.1 we present a short account of the available techniques and the definitions on which they are based.

4.1.1 Related Work

Statistical measures such as the auto-correlation function (Peebles 1980) of the matter distribution in the web have been the mainstay of cosmological studies over many decades. However, while this second-order measure of clustering does not contain any phase information (one may e.g. always reproduce a distribution with the same 2nd order moments and random Fourier phases), the auto-correlation function is not sensitive to the existence of complex spatial patterns. Higher order correlation functions only contain a very limited amount of such structural information, while in practical observational circumstances, it quickly becomes cumbersome to measure them as the magnitude of the error increases drastically with increasing order.

The first attempts towards characterizing complex geometric patterns in the galaxy distribution mainly involved heuristic measures. Early examples of techniques addressing the global connectivity of structure in the Universe are percolation analysis (Shandarin & Zeldovich 1983) and the minimum spanning tree of the spatial galaxy

distribution (Barrow et al. 1985; Alpaslan et al. 2014). While these are useful global descriptions, they do not capture and describe local characteristics of the mass distribution.

More elaborate and advanced techniques have been developed in recent years. Several of these methods apply sophisticated mathematical and visualization techniques, involving geometric and topological properties of the cosmic mass distribution. There are a multitude of different methods for detecting filaments, based on a range of different techniques. We may recognize several categories of techniques.

One class of methods seeks to describe the local geometry on the basis of the Hessian of the density field (Aragón-Calvo et al. 2007a,b; Bond et al. 2010; Cautun et al. 2013) or closely related quantities such as the tidal force field (Hahn et al. 2007; Forero-Romero et al. 2009) or the velocity shear field (Libeskind et al. 2012; Cautun et al. 2013). The Hessian provides direct information on the local shape and dynamical impact of the corresponding field. The morphological elements of the cosmic web are identified by connecting the areas within which a specific range of anisotropies is registered.

These studies concentrate on a single scale by appropriately smoothing the field, and do not consider the multi-scale nature of the cosmic mass distribution. The Hessian based Nexus/MMF technique (Aragón-Calvo et al. 2007b), which was perfected into a versatile and parameter-free method (Cautun et al. 2013), implicitly takes into account the multi-scale nature of the web-related fields. It accomplishes this by a scale-space analysis of the fields. At each location the optimal morphological signal is extracted via the application of a sophisticated filter bank applied to the Hessian of the corresponding fields in scale space. The application of this machinery has enabled thorough studies of the hierarchical evolution and buildup of the cosmic web (Cautun et al. 2014).

A promising and highly interesting recent development has opened up the path towards dynamical analysis of the evolving mass distribution in full six-dimensional phase-space (in which the position of each mass element is specified by its space coordinates and velocity/momentum). In the 6D phase space, the cosmic mass distributions defines a 3D sheet. Independently, three groups arrived at tessellation based formalisms that exploit the evolving structure and folding of the *phase space sheet* in phase space (Shandarin 2011; Abel et al. 2012; Neyrinck 2012) (also see e.g. Shandarin et al. (2012)). The number of folds of the phase space sheet at a given location indicates the number of local velocity streams, and forms a direct indication of the morphology of the local structure. Interestingly, the resulting characterization of the web-like distribution, the Origami formalism of Neyrinck (2012) for example, appears to resemble that of the Nexus/MMF formalism (Cautun et al. 2014).

An entirely different class of techniques, based on statistical methods, have also been used to recover the filamentary patterns in the Universe. The key idea behind these techniques is to treat the galaxy distribution as a Markov point process. Within this class, particularly worth mentioning is the Bisous or the Candy model, which has been used by Stoica et al. (2007) to detect the filamentary network [2]. In

[2]See Stoica et al. (2005), for a detailed description of this object point process which is used to characterize an observed point distribution by a fitting procedure using global optimization techniques.

this model, one places a random configuration of interacting geometric cylindrical objects on the point process to detect filamentary structure. It has been developed into a versatile, statistically solid yet computationally challenging formalism for the identification of filaments in a spatial point distribution, such as N-body simulations and galaxy redshift surveys (Tempel et al. 2013). An additional example of a method involving statistical analysis of a geometric model is that of Genovese et al. (2010), which seeks to describe the filamentary patterns of the cosmic web in a non-parametric way by recovering the medial axis (Blum 1967) of the point-set of galaxies.

The fourth major class of methods, the one which we will also pursue in this paper, exploits the topological structure of the cosmic mass distribution. The fundamental basis of these methods is Morse theory (Milnor 1963). The geometric structure of the *Morse-Smale complex* (Edelsbrunner et al. 2001) naturally delineates the various morphological components on the basis of the connections between the critical points of the density fields and the higher dimensional cells that are incident on the critical points. Various Morse theory based formalisms have been applied to the identification of components of the cosmic web. One of the first applications concerned the detection of voids in the cosmic density field. The Watershed Void Finder (Platen et al. 2007) identifies these with the watershed basins around the density minima. The SpineWeb procedure (Aragón-Calvo et al. 2010) extended the watershed transform towards the detection of the full array of structural components, filaments, walls and voids. These techniques use a user-defined filter to incorporate the multi-scale structure of the cosmic density field.

A natural topological means to address the multi-scale topological structure emanates from the concept of persistence (Edelsbrunner et al. 2002). It provides a natural recipe for detecting and quantifying the components of the cosmic web in a truly hierarchical fashion. Sousbie (2011) and Sousbie et al. (2011) has exploited and framed this in an elegant and impressive framework, the DisPerSE formalism. Following the construction of the Morse-Smale complex, they proceed to simplify it. The simplification proceeds by canceling pairs of critical points iteratively, where each pair represents a structure in the cosmic web. Topological persistence is invoked to order the critical point pairs. However, this measure of importance is not unique, and one may consider alternatives, dependent on the specific interest and purpose.

In effect, to tackle similar issues in other visualization areas, a range of variations have been proposed in other studies (Günther et al. 2012; Reininghaus et al. 2011; Weinkauf & Günther 2009). Weinkauf & Günther (2009) describe the concept of separatrix persistence, where they compute the strength of separation of points on a separatrix curve (in 2D) connected to a saddle as the sum of the absolute differences of function values of the saddle and the extrema connected to it. This concept is extended to 3D separating sheets by Günther et al. (2012). Reininghaus et al. (2011) develop the concept of scale-space persistence where they accumulate the absolute difference in function value measure of critical points across a hierarchy of derived functions. The set of derived functions are generated by smoothing the function using a family of Gaussian kernels of increasing variances. This is similar to the Multi-

scale Morphology Filter Nexus/MMF (Aragón-Calvo et al. 2007b; Cautun et al. 2013) described above. Both methods adopt the scale space formalism as the first step to detect features at multiple scales. However, scale space persistence and separatrix persistence, disregard specific density regimes of interest and are potentially inappropriate when small scale features with specific density characteristics are of interest.

4.1.2 Present study: contributions

In the present study, we describe and introduce a technique for the identification of filaments based on the topological characteristics of the density field. A key aspect of the proposed technique is its interactive nature, involving a tunable density parameter. Specifically, we describe the following contributions:

- We describe Felix[3]: a topology based framework for visual exploration of filaments in the cosmic web. In particular, we develop a query framework to extract filamentary structures from a hierarchy of Morse-Smale complexes of the density field. The filaments in Felix are parameterized by the density values of the maxima and the 2-saddles that define them.
- Using Felix, we develop a semi-automatic structure finder that classifies galaxies as cluster/filamentary or not. We demonstrate its efficiency through two tests. First, using the Voronoi Kinematic model as a benchmark, we demonstrate that we are able to recover the classification with high efficiency. Second, we show that the classifications are quantitatively comparable to, and in several cases better than, existing classifiers.
- We investigate the nature of filaments in two different density regimes from the ΛCDM simulations. The first concerns filaments in the high density regions around compact dense clusters, which are known to function as the transport channels along which matter moves into the clusters. A second regime concerns the tenuous low-density filaments found in low-density void regions. We present an additional experiment, where we investigate the nature of three classes of filaments in a relatively cleaner region of a ΛCDM dataset.
- We describe an efficient structure based volume rendering enhancement routine that allows us to highlight the density distribution in regions that are close to the selected features.

The distinction between noise and significant structures is often ill-defined, and at occasions noise may be confused with genuine structures in the hierarchically evolved mass distribution (see Figure 4.4 and the caption thereof for an illustration). This problem is more pronounced when one studies the properties of tenuous filaments and walls in low density void-like regions. For the understanding of the formation and evolution of galaxies in such regions, we need to assess the possible dependence of galaxy and halo properties on the morphology and density of the local environment. This must be based on the successful extraction of filaments in low

[3]The name Felix is formed from an abbreviation of *Fil*ament *ex*plorer.

density regions and the correct identification of galaxies associated with them. In view of this, we include an interactive handle on the density regimes so that one can concentrate on and probe structures in specific density regimes.

The remainder of the paper is organized as follows. Section 4.2 introduces necessary background material. Section 4.3 describes Felix. Section 4.4 introduces the cosmological datasets used in the experiments. Section 4.5 discusses the application of Felix as a structure finder and as a tool for exploring filaments in different density regimes. Section 4.6 discusses an application based on Felix for a volumetric enhancement routine. This is used to supress the opacity of the density field far removed from the structures of interest, thereby reducing the visual clutter. Section 4.7 concludes the paper by summarizing the main results and possible future directions.

4.2 Background

This section reviews relevant background on Morse functions, the Morse-Smale complex, and topological simplification. This is a necessary prerequisite for understanding the definition of filaments and extraction methods described in the subsequent sections.

4.2.1 Morse theory and the Morse-Smale complex

Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be a real-valued scalar function defined on a manifold $\mathbb{M}$. Critical points of f are points of f where the gradient of f vanishes, i.e., $\nabla f = 0$. Morse theory is the study of the relationship between the topology of level sets of scalar functions and the critical points of the function. The function f is said to be a *Morse function* if all of its critical points are non-degenerate, i.e., the *Hessian* of f, equal to the matrix of second order partial derivatives, is non-singular. The non-degeneracy condition imposes a locally quadratic form for f within a small neighborhood of its critical points. In other words, using a coordinate transformation, the function near a critical point p of the n-dimensional manifold $\mathbb{M}$ can be written as $f_p(x) = f(p) \pm x_1^2 \pm x_2^2 \pm ... \pm x_n^2$. The *index* of p is equal to the number of negative quadratic terms in the above expression. In 3D, the index 0 corresponds to minima, the index 1 corresponds to 1-saddles, the index 2 corresponds to 2-saddles, and the index 3 corresponds to maxima. An *integral line* is a maximal curve in the domain, whose tangent aligns with the gradient of f at every point. The function f increases along the integral line and its limit points are the critical points of f.

The set of all integral lines that originate at the critical point p together with p is called the *ascending manifold* of p. Similarly, the set of all integral lines that terminate at the critical point p together with p is called the *descending manifold* of p. The ascending manifolds (similarly, the descending manifolds) of all critical points partition the domain. The ascending manifold of a critical point with index i has dimension $n - i$, where n is the dimension of the domain. Thus, the ascending manifold of a minimum is a three dimensional cell, the ascending manifold of a 1-saddle is a two dimensional sheet, the ascending manifold of a 2-saddle is a one dimensional arc (see Figure 4.1), and the ascending manifold of a maximum is equal to the maximum. The converse is true for the descending manifold i.e., the descending

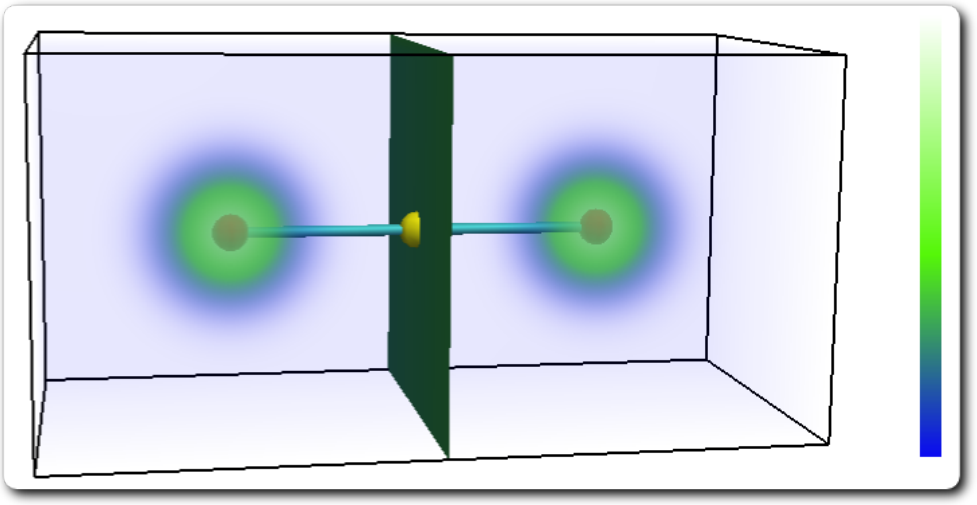


Figure 4.1 Ascending manifolds of a 2-saddle (yellow sphere). The scalar function is a sum of two 3D Gaussians centered on either side of the volume. The two arcs incident on the 2-saddle constitute the ascending manifold and terminate at the two maxima (red spheres) of the scalar function.

manifold of a critical point with index i has dimension i. The *Morse-Smale complex* is a partition of the domain into cells formed by the collection of integral lines that share a common source and a common destination. The function f is called a *Morse-Smale function* if the ascending and descending manifolds of all pairs of critical points intersect only transversally i.e., if the ascending and descending manifolds of two critical points intersect, then the intersection has dimension exactly equal to the difference in the indices of the two critical points. The critical points, referred to as *nodes*, together with the 1-manifolds that connect them, referred to as *arcs*, form the 1-skeleton of the Morse-Smale complex, which is referred to as the *combinatorial structure* of the Morse-Smale complex.

4.2.2 Morse-Smale complex simplification

The Morse-Smale complex may be simplified by repeated application of the *topological cancellation* procedure. Topological cancellation eliminates a pair of critical points in the Morse-Smale complex connected by a single arc. The resulting complex is representative of a smoother version of f, which may be obtained by a local smoothing operation within an infinitesimal neighborhood of the arc. The canceled pair of points are no longer critical in the smoother version of f. By repeatedly applying topological cancellations, one obtains a simpler Morse-Smale complex, where undesirable pairs of critical points are eliminated. This simplified Morse-Smale complex is representative of the function f with several smoothing operations applied to it. Pairs of critical points that are multiply connected in the Morse-Smale complex, referred to as *strangulations*, cannot be simplified by topological cancellation.

The cancellation procedure mandates changes to the combinatorial structure and ascending/descending manifolds of critical points that survive the cancellation. These

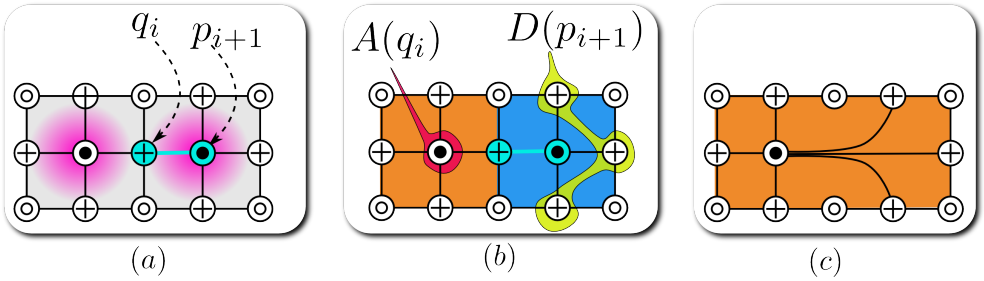


Figure 4.2 Topological cancellation of a pair of critical points in a 2D Morse-Smale complex. (a) Morse-Smale complex of the function shown in Figure 1 restricted to a 2D slice. Maxima are denoted by $\odot$, saddles by $\oplus$, and minima by $\circledcirc$. A pair (p_{i+1}, q_i) of critical points connected by a single arc is scheduled to be canceled. (b) $D(p_{i+1})$ is the set of surviving index i critical points connected to p_{i+1} and $A(q_i)$ is the set of surviving index $i+1$ critical points connected to q_i. (c) Combinatorial realization: connect all critical points $D(p_{i+1})$ to those in $A(q_i)$. Geometric realization: merge the descending manifold of p_{i+1} with those of critical points in $A(q_i)$. Merge the ascending manifold of q_i with those of critical points in $D(p_{i+1})$.

changes are respectively referred to as the combinatorial and geometric realization of the cancellation. Let p_{i+1} and q_i denote the pair of critical points to be eliminated with $i+1$ and i being the respective Morse indices. The combinatorial realization proceeds by first removing p_{i+1} and q_i as well as all arcs incident upon them. Next, new arcs are introduced between every surviving index i critical point that was connected to p_{i+1} and every surviving index $i+1$ critical point that was connected to q_i. The geometric realization merges the descending manifold of p_{i+1} with the descending manifold of each surviving index $i+1$ critical point connected to q_i. Analogously, the ascending manifold of q_i is merged with the ascending manifold of each surviving index i critical point connected to p_{i+1}. Figure 4.2 illustrates both realizations of a topological cancellation procedure applied to the Morse-Smale complex of a two-dimensional slice of the function shown in Figure 4.1.

The ordering of cancellation pairs plays a crucial role in determining the resulting structure of the Morse-Smale complex and its geometry. To simplify the Morse-Smale complex, pairs of singularly connected critical points having the least absolute difference in function value are iteratively canceled. This approach is equivalent to the notion of topological persistence (Edelsbrunner et al. 2001, 2002) for 2D Morse-Smale complexes, but not necessarily for 3D Morse-Smale complexes (Günther et al. 2014).

4.2.3 The hierarchical Morse-Smale complex

A sequence of cancellations results in a hierarchical sequence of Morse-Smale complexes $MSC_0, MSC_1, ..., MSC_n$, where each Morse-Smale complex is a simpler version of the preceding Morse-Smale complex containing fewer critical points. Morse-Smale complex MSC_i is said to be coarser than MSC_j if $i > j$ and finer if $i < j$. The version index i enumerates the Morse-Smale complexes in the hierarchy. Each non-zero version of the Morse-Smale complex, MSC_i, is associated with the absolute difference in function value, t_i, of the pair of critical points canceled in the preceding version, MSC_{i-1}. As each iteration selects the pair of critical points with the least ab-

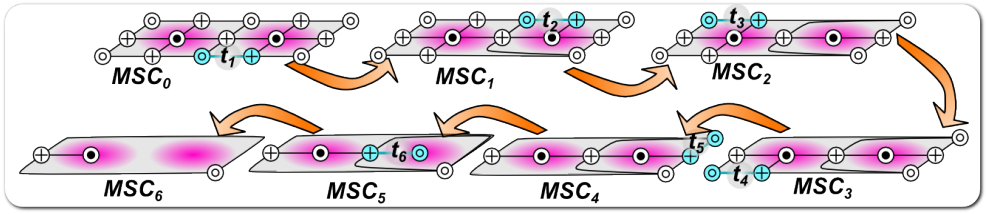


Figure 4.3 Hierarchical Morse-Smale complex. A family of Morse-Smale complexes generated by iteratively canceling pairs of critical points. MSC_0 is the Morse-Smale complex of a 2D equivalent of the function shown in Figure 4.1. It is simplified to generate a coarser version, MSC_1, by canceling a pair of critical points (cyan) connected by a single arc and having least absolute difference in function value t_1. Successive versions MSC_i are computed similarly by selecting arcs so that $t_0 = 0 \leq t_1 \leq \ldots \leq t_6$.

solute difference in function value, the sequence of t_i's is monotonically increasing i.e. $(t_0 = 0) \leq t_1 \leq t_2 \ldots \leq t_n$. For completeness of the sequence, t_0 is set to zero. Figure 4.3 illustrates a hierarchy of Morse-Smale complexes of a 2D equivalent of the function shown in Figure 4.1.

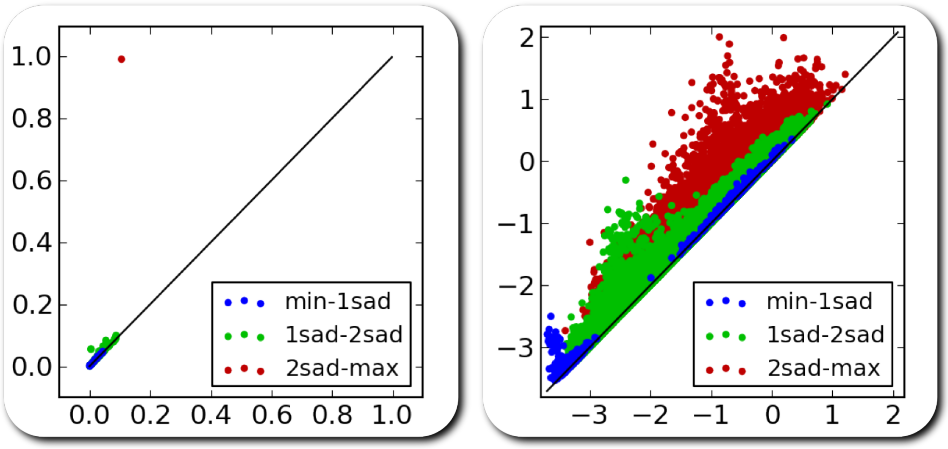
It is not necessary to explicitly store all versions of the hierarchy of Morse-Smale complexes. Instead, the combinatorial representation of only MSC_0 is computed initially. Subsequently, MSC_i can be obtained from a finer MSC_{i-1} by performing topological cancellation (see Section 4.2.2). Analogously, MSC_i is obtained from a coarser MSC_{i+1} by applying the inverse operation of a topological cancellation.

4.3 Methodology

Exploring the filamentary patterns of the cosmic web is challenging because of the large range of the spatial scales and density range it exhibits. A proper characterization should also account for the hierarchical nature of structures, which adds considerable challenges to the task. Though there exist different notions of filaments, the primary evidence relied upon for extraction and analysis is most often visual. It is therefore not surprising that structure finding methods often visually verify results by superimposing the extracted structures upon visualizations of the density field or the particle distribution. However the visualization plays a role only after structure extraction process in these methods. We differ in this respect by providing the capability to interact with the structure finding procedure and extract structures that are visually relevant. To accomplish such a visual exploration framework, a succinct model of filament definition, an efficient representation of hierarchical structures, and an appropriate query mechanism that supports the extraction of these structures are paramount. The following exposition details our framework on these terms.

4.3.1 Density estimation and filament modeling

Cosmological simulations are N-body particle experiments that simulate structure formation and evolution by tracing positions of the particles under the influence of physical laws. In the observational reality, the information about structures in the cosmos comes through observing the galaxies. The galaxies can be treated as particles also for the purpose of analysis in the context of large scale structures.



(a) (b)

Figure 4.4 (a) A scatter plot of the function values of the canceled critical point pairs for the function shown in Figure 4.1. A 2-saddle-maximum pair is the only pair that is far removed from the diagonal. This corresponds to cancellation of the 2-saddle with a maximum that represents one of the Gaussians in Figure 4.1. Other pairs close to the diagonal represent insignificant features that manifest due to the added Gaussian noise as well as sampling noise. (b) A scatter plot of the function values of the canceled critical point pairs for the Voronoi-Kinematic dataset B (see Section 4.4.1). No discernible separation of points is seen, though there are many points that are far removed from the diagonal. Thus, no clear global simplification threshold may be used for filament extraction.

The input to Felix is the logarithm of the density scalar field on the given 3D domain of interest. The domain could be 3D structured grids or tetrahedral meshes, with the density specified on the vertices of the grid/mesh. We find that the logarithm of the density field, instead of the density field itself, resolves the structures with more visual clarity. This has also been independently established in an earlier work (Cautun et al. 2013). Additionally, the input may be specified as a distribution of particles within a 3D region of interest. This could be a snapshot from a cosmological simulation, or galaxies in real observational data. We use the Delaunay tessellation field estimator (DTFE) (Schaap & van de Weygaert 2000a; van de Weygaert & Schaap 2009a) to estimate the density of the input particles in the 3D region of interest. This procedure begins by computing the periodic Delaunay triangulation on the points (simulation particles or galaxies). Next, the density at each vertex of the triangulation is estimated by the inverse of the volume of the tetrahedra incident upon it. Finally, the density is linearly interpolated onto the edges, faces, and tetrahedra of the Delaunay triangulation to yield a piecewise linear density function on the domain. The periodic Delaunay triangulation, computed by the DTFE procedure, is used to represent the domain.

The Morse-Smale complex of the logarithm of the density field is computed. Filaments are modeled as the ascending manifolds of 2-saddles of the Morse-Smale complex. These arcs represent paths of steepest descent from the two maxima merging at the 2-saddle. This 2-saddle represents the lowest density point along the arcs connecting the two maxima. A schematic illustration of this is presented in

Figure 4.1. There are many algorithms available in the literature to compute the 3D Morse-Smale complex. The algorithms are primarily based on either the quasi Morse-Smale complex formulation (Gyulassy et al. 2007; Edelsbrunner et al. 2003) or Forman's (Forman 2002) discrete Morse theory (Gyulassy et al. 2008; Robins et al. 2011; Shivashankar et al. 2012; Shivashankar & Natarajan 2012). We use a parallel algorithm based on the latter approach (Shivashankar & Natarajan 2012) resulting in fast computation even for large datasets.

The density field is rarely smooth and several local maxima obscure a view of the larger scale behavior of the density field. This is especially true if the density field is computed on the raw particle distribution, where the density field tends to be spiky and with a lot of fluctuations in the high density cluster-like regions. The Morse-Smale complex is simplified by iteratively canceling pairs of singularly connected critical points with least absolute difference in function value to generate a hierarchy of Morse-Smale complexes.

In most applications, a specific version of the Morse-Smale complex from the hierarchy is chosen based on a perceptibly clear separation of noise and features. One way to choose such a threshold separating noise and feature is by using a scatter plot of the function values of canceled critical point pairs (see Figure 4.4) where the lower function value among the pair corresponds to the x-coordinate and the higher function value corresponds to the y-coordinate. In datasets where topological features are well separated (see Figure 4.4a), pairs representing significant features appear far away and isolated from the diagonal. In such cases, the coarsest Morse-Smale complex version wherein the insignificant pairs are removed is selected for feature analysis/extraction. However, this strategy is not easily applicable to cosmology datasets (see Figure 4.4b). A well defined separation is rarely discernible, though there are many scatter points that are far removed from the diagonal. Hence, we drop the assumption that we must work with a specific version of the Morse-Smale complex. Instead, we query for features across all Morse-Smale complexes in the hierarchy, as discussed in the following sub-section.

4.3.2 Density range based filament selection

Cosmic filaments exhibit a large range of variation in their density characteristics. Indeed, one expects filaments to be present both in void like regions and between cluster like regions. While strong dense filaments in between clusters define the spine of the Cosmic Web, in the hierarchically evolving mass distribution we encounter a wide spectrum of ever more tenuous filaments on smaller mass scales. Small filaments define the directions of mass inflow into galaxies, and form a crucial component in the formation of galaxies (Codis et al. 2012). Even more tenuous are the systems of filaments stretching over the hollows of voids, often conspicuously aligned along the direction defined by neighbouring superstructures. The understanding of this network is tightly related to the issue of the "missing" dwarf galaxies in voids (Peebles 2001). While this illustrates the complexity of the multiscale filigree of filaments in the Cosmic Web, we follow a strategy in which we focus our attention on specific aspects and details of the cosmic web. Dependent on the identity of objects and structures of interest, we wish to be able to zoom in on to the correspond-

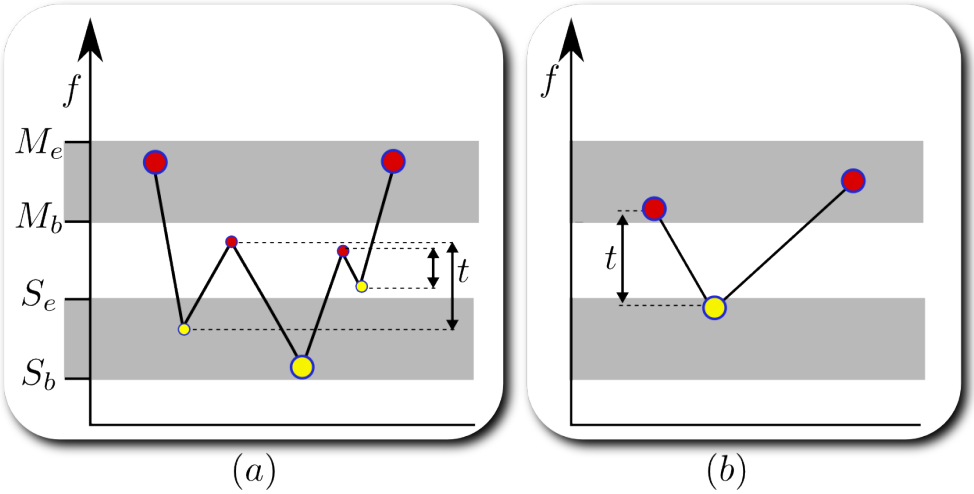


Figure 4.5 (a) Filaments are modeled as the ascending paths of 2-saddles connecting two extrema. The 2-saddles are filtered based on the range constraints $[M_b, M_e]$ and $[S_b, S_e]$ on the highest and lowest values respectively along the ascending paths. The highest values along the 2-saddle's ascending manifold are at extrema and the lowest value is at the 2-saddle. The function along the paths needs to be simplified as it is rarely smooth. In the illustration, a simplification threshold of t reveals a filament with appropriate density characteristics. However, imposing such a threshold uniformly will cause another filament (b) having the required density characteristics to be destroyed. It is therefore necessary to extract filaments by querying all Morse-Smale complexes within a given hierarchy.

ing filamentary network. This is largely dependent on the mass scales of the objects involved, and the density values of the corresponding filament generating density peaks (Aragón-Calvo et al. 2010; Cautun et al. 2014).

Following this rationale, we translate this strategy into the use of queries that depend on the density properties of interest. Specifically, we query for filaments by specifying the density range $[M_b, M_e]$ of the clusters they connect (the maxima at the end points), as well as the density range $[S_b, S_e]$ of the lowest point along the connecting path (the density range of the 2-saddles). Figure 4.5 conceptually illustrates the characterization of filaments using density ranges, where density along filaments varies significantly necessitating simplification.

Algorithm 1 lists the algorithm to process such a query. The algorithm accepts, together with the combinatorial Morse-Smale complex MSC, the density ranges of 2-saddles $[S_b, S_e]$ and maxima $[M_b, M_e]$ as input (the subscripts $_b$ and $_e$ denote the beginning and ending of each density range respectively). The algorithm returns a list of 2-saddles that satisfy the above criteria together with the maximal Morse-Smale complex version in which they do so.

The algorithm begins by creating a list S of 2-saddles that have their function value in the given 2-saddle range $[S_b, S_e]$. Then, for each 2-saddle in S, a Morse-Smale complex version in which it possibly connects two maxima within $[M_b, M_e]$ is computed. The appropriate version is given by the minimum of three version indices v_a, v_b and v_c.

Algorithm 1 ($[S_b, S_e]$, $[M_b, M_e]$)

SELECT2SADDLES

1: $Sver$ = Empty Map
2: $S = \{s | s$ is a 2-saddle,$S_b < f(s) < S_e\}$
3: **for all** $s \in S$ **do**
4: $v_a = Max\{i | s$ is not simplified in $MSC_i\}$
5: $v_b = Max\{i | t_i < M_b - f(s)\}$
6: $v_c = Max\{i | s$ connects distinct Maxima in $MSC_i\}$
7: $Sver[s] = Min(v_a, v_b, v_c)$
8: **end for**
9: Sort S by Sver
10: $Ssel$ = Empty Set
11: **for all** $s \in S$ **do**
12: $i = Sver[s]$
13: $m_a, m_b =$ Maxima connected to s in MSC_i
14: **if** $M_b < f(m_a), f(m_b) < M_e$ **then**
15: insert $(s, Sver[s])$ in $Ssel$
16: **end if**
17: **end for**
18: **return** $Ssel$

The version index v_a is the finest Morse-Smale complex version in which the 2-saddle s survives. In other words, s is canceled in MSC_{v_a+1} but not in MSC_{v_a}. This is pre-computed by examining the cancellation sequence. The version index v_b corresponds to the last Morse-Smale complex version at which the 2-saddle s connects two maxima, both with function value less than M_e. This is possible because in successive versions of the Morse-Smale complex, the maxima connected to a 2-saddle via the same arc form an increasing sequence in terms of their function value. Thus, in the version where the absolute difference in function value of the last canceled pair is less than $M_b - f(s)$, the 2-saddle s still possibly connects two maxima with function value less than M_e. The version index v_c is the last Morse-Smale complex version at which the 2-saddle s separates distinct maxima. In other words, it is not a strangulation in MSC_{v_c}. As a consequence of the cancellation preconditions, once a strangulation is created by a 2-saddle, it may be destroyed only by canceling the 2-saddle with a 1-saddle. Thus there exists a maximal version index v_c after which the 2-saddle remains connected to a single maximum. The version index v_c is -1 when the 2-saddle is a strangulation in the initial Morse-Smale complex. In this case, the 2-saddle is not considered in further steps and is removed from S. Again, this is easily pre-computed for each 2-saddle by examining the cancellation sequence.

The 2-saddles in the set S are sorted based on their version indices. This is done to optimize switching between the required Morse-Smale complex versions. Next, each 2-saddle s in S is checked to see if it separates two maxima within the maxima density range $[M_b, M_e]$. The list of 2-saddles that fulfill all of the above criteria is returned together with the associated version number of each 2-saddle. This above

list of 2-saddles is used to extract the filament geometry. Specifically, the ascending manifold of each 2-saddle is extracted from the corresponding version of the Morse-Smale complex. This may be done efficiently using the *cancellation merge DAG* data structure discussed by (Gyulassy et al. 2012).

In some situations, it is desirable to perform some simplification to eliminate Poisson noise introduced due to meshing the domain. In these cases, a global simplification specifically for noise elimination, can be optionally introduced. Specifically, Algorithm 1 returns only those 2-saddles that survive in hierarchical Morse-Smale complex versions above a specified threshold T_s, where T_s is specified as a normalized fraction of the range of log-density values (normalized to $[0, 1]$). T_s is set to 0.0 unless specifically mentioned. Similar to the inputs of Algorithm 1, T_s may be updated during run-time.

4.4 Model Description

In this section, we briefly describe the models used to test our filament detection routine. These are the Voronoi evolution models and ΛCDM cosmological simulations.

The Voronoi models provide us with vital quantitative information on the sensitivity of Felix to anisotropic filamentary patterns in the galaxy distribution. To this end it is of key importance that the Voronoi models have an a-priori known population fraction in different morphological elements: clusters, filaments, walls and voids. This makes them perfect test models for evaluating success and failure rates of the various identification methods.

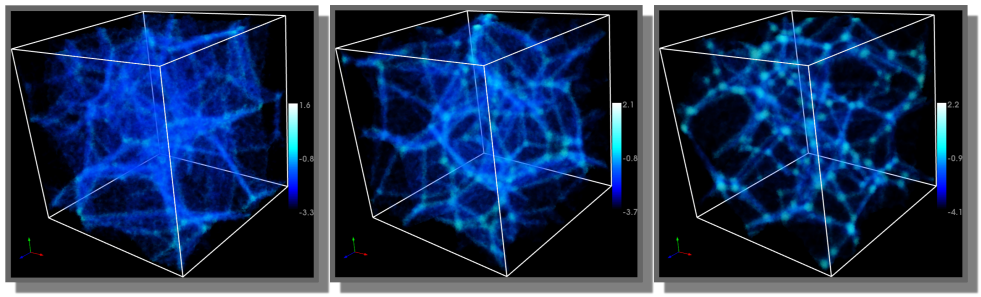
Although they involve filaments with a broad distribution of densities, the Voronoi models do not incorporate the multi-scale web-like patterns seen in realistic cosmological scenarios. To assess this aspect of the cosmic mass distribution, we turn to simulations of structure formation in the standard ΛCDM cosmology. Implicitly these include all relevant physical and dynamical processes of the evolving cosmic dark matter distribution. However, as we have no control over all aspects of the emerging mass distribution in ΛCDM simulations, for testing purposes they are not as informative as Voronoi models [4]

4.4.1 Voronoi evolution models

The Voronoi evolution models are a class of heuristic models for cellular distributions of galaxies that mimics the evolution of the Megaparsec universe towards a weblike pattern. They use Voronoi tessellations as a template for distribution of matter and related galaxy population (van de Weygaert & Icke 1989; van de Weygaert 1991; Weygaert 2007), and its subsequent evolution.

In these models, one begins by fixing an underlying Voronoi skeleton, defined by a small set of randomly distributed nuclei in the simulation box. One then superposes a set of N randomly distributed particles on this skeleton. The resulting spatial

[4] In the literature, several studies use simplistic models using Voronoi tessellations. The models we use here are considerably more sophisticated, and represent a rather realistic depiction of the cosmic web in void-dominated cosmologies, (see e.g. van de Weygaert & van Kampen (1993); Sheth & van de Weygaert (2004); Weygaert (2007); Aragon-Calvo & Szalay (2013)).



(a) (b) (c)

Figure 4.6 Density rendering of the snapshots in the Voronoi evolution time-series: (a) is the least evolved stage with almost 50 percent particles located in the Voronoi cells, (b) at a medium stage of evolution, while (c) is the most evolved stage of the model.

distribution of particles in the model is obtained by projecting the initially random distribution of particles on to the faces, edges, and nodes of the Voronoi tessellation. This results in a pattern in which one can distinguish four structural components: *field* particles located in the interior of Voronoi cells, *wall* particles within and around the Voronoi faces, *filament* particles within and around the Voronoi edges and *cluster* particles within and around the Voronoi nodes.

One particular class of Voronoi clustering models are the Voronoi kinematic models, which seek to approximate the dynamical evolution of the large scale cosmic mass distribution. These models involve a continuous flow of galaxies towards the nearest wall, along a filament at the wall's edge, and subsequently towards the final destination, a vertex of the Voronoi tessellation. This motion is regulated by the increase of mean distance between the galaxies, an expression of void expansion and evacuation as a function of time.

The Voronoi models used in our experiments have 262,144 particles distributed along the vertices, edges, faces and cells of the Voronoi skeleton in a box of side-length $200h^{-1}$Mpc. The skeleton is generated by 32 randomly placed nuclei in the box. For the least evolved stage, most of the particles are in the cells, while for the most evolved stage, most of the particles are located in and around clusters. Table 4.1 presents the percentage distribution of particles in the various structural elements, as it changes with time. Stage 1 (dataset A) is the least evolved, while Stage 3 (dataset C) is the most evolved. The particles in and around the nodes, edges, and walls are Gaussian distributed around these elements, characterized by a thickness scale R_f which is the standard deviation of the distribution. For our models, $R_f = 2h^{-1}$Mpc. Figure 4.6 shows the density rendering of the Voronoi kinematic datasets used for evaluating Felix.

4.4.2 ΛCDM cosmological simulations

The ΛCDM simulations are fully physical models that trace the distribution and evolution of dark matter in the universe based on current understanding of real physical

	cell	wall	filament	cluster
A	29.88%	43.57%	22.20%	4.33%
B	9.82%	32.13%	38.62%	19.42%
C	3.5%	16.50%	28.70%	51.30%

Table 4.1 The relative abundance of particles in each structural element throughout the course of evolution.

laws. Dark matter is the gravitationally dominant matter component in the Universe and constitutes the major fraction of matter. As it is known to only interact gravitationally, modeling the behavior of dark matter is computationally fast and efficient. Such dark matter simulations form one of the principal tools towards understanding the evolution of the matter distribution in the Universe.

The cosmological simulations that we used follow the standard ΛCDM cosmology. In this model, the matter content of this Universe is dominated by collisionless *cold dark matter* (CDM) particles. The biggest contribution to the energy content of this Universe comes from dark energy, in the form of the cosmological constant Λ (see (Komatsu et al. 2009)), which drives its accelerated expansion at the current epoch.

To present the results of our visual exploration framework, we use the Cosmogrid simulations (Ishiyama et al. 2013). It is a suite of simulations in a box of size $21h^{-1}Mpc$, each differing in the number of particles. The particular simulation we use for our study comprises of 512^3 particles. This is a relatively small scale in the context of the cosmic web. The mass resolution achieved is $8.21 \times 10^6 M\odot$. The initial conditions are setup at $z = 65$ using the Zel'dovich approximation (Zel'dovich 1970). The log-density field is available on a $128 \times 128 \times 128$ structured grid.

Particularly characteristic in the evolving mass distribution of the Cosmogrid simulation is the large central under-density, surrounded by a range of smaller voids near its outer edge. In combination with its extremely high spatial resolution and state-of-the-art dynamic range, this renders the Cosmogrid simulation uniquely suited as a testbed for a case study of the internal structure of voids. It was precisely this circumstance that formed the rationale behind its exploitation in a previous study of the formation of dark halos along tenuous void filaments (Rieder et al. 2013).

4.5 Results and Discussion

In this section, we demonstrate and discuss the salient features and potential applications of Felix. First, we evaluate the filaments extracted using Felix and compare with those extracted using MMF, SpineWeb, and DisPerSE using the Voronoi kinematic datasets. Next, we present a visual exploration of the filaments in the Cosmogrid simulation. The methods described above were implemented and tested on a computer system with an Intel Xeon(R) 2.0 GHz CPU and 8GB of RAM.

4.5.1 Filaments in the Voronoi model: a comparison

Here we present an analysis of the filamentary structures extracted using Felix, and compare the results with the techniques detailed above. The comparison study concerns the analysis results obtained for the set of heuristic Voronoi evolution models described above. Since they are input parameters, in these models the classification of galaxies as void, wall, filament, and cluster are known a-priori. Following the application of one of the detection techniques we may then examine the validity and authenticity of the extracted structures by direct comparison with the true identity of a galaxy.

For the comparison study, we define two measures. One quantifies the true detection rate of a method, the other the false identifications. We classify all galaxies within a distance d from the extracted structures to be filament and cluster particles and the others to be void and wall galaxies. For a given set of structures extracted from a given dataset and a distance d, the true positive classification rate TP_d is defined as

$$TP_d = \frac{\text{\# filament and cluster galaxies correctly classified}}{\text{\# filament and cluster galaxies}}.$$

Similarly, the false positive classification rate FP_d is defined as

$$FP_d = \frac{\text{\# filament and cluster galaxies incorrectly classified}}{\text{\# filament and cluster galaxies}}.$$

A large separation between these two measures indicates good discriminatory power of the classifier, and thus the proximity of relevant galaxies to the extracted structures.

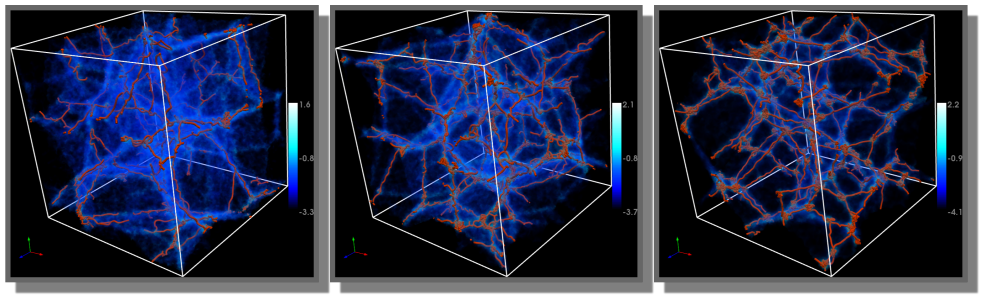
As we discuss in more detail below, the Felix's true and false detection rates are comparable, and in some situations better, than those obtained by DisPerSE, SpineWeb and Nexus/MMF. A brief description of SpineWeb, and Nexus/MMF is also provided. As the DisPerSE methodology is closely related to Felix, in the following paragraphs, we briefly describe it and contrast it against Felix.

Felix

Figure 4.7 shows the extracted filaments for the Voronoi kinematic datasets A, B, and C using Felix. The input density range parameters for Algorithm 1 are selected interactively, using the visualization information of the procedure. Each update is accomplished within 2-3 seconds. This enables interactive visual feedback so that parameters may be adjusted further in subsequent iterative steps.

Comparison A: Felix and DisPerSE

DisPerSE (Sousbie 2011; Sousbie et al. 2011) is a closely related structure finder that also uses the Morse-Smale complex of the logarithm of the density field. It simplifies the Morse-Smale complex using Topological Persistence. The function values are normalized by the rms of the density field fluctuation with respect to the mean, and the significance level for simplification is quoted in this unit. Felix is closely related to DisPerSE as both use the Morse-Smale complex of the log-density



(a) (b) (c)

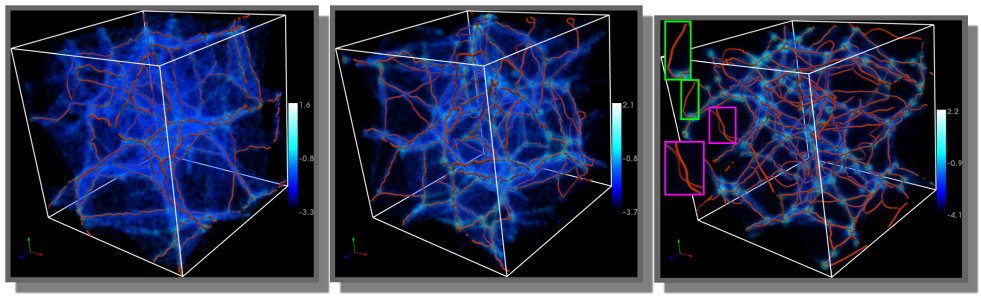
Figure 4.7 Filamentary structures extracted from datasets A, B, and C of the Voronoi evolution time-series using Felix with parameters for Algorithm 1 set as follows: (a) $[S_b, S_e] = [10^{-1.6}, \infty]$ and $[M_b, M_e] = [10^{-0.1}, \infty]$, (b,c) $[S_b, S_e] = [10^{-1.6}, \infty]$ and $[M_b, M_e] = [10^{0}, \infty]$. Filaments are shown as orange tubes along with a volume rendering of the log-density field. The dense knot like structures show filaments within cluster-like regions.

field and involve feature extraction from it. DisPerSE defines significant features as only those that remain unsimplified using the user defined significance threshold. It ignores the density range characteristics of the extracted features. A significant consequence is that filaments within void-like regions and cluster like regions are ignored/simplified away. If they are retained, then the mixing of features causes visual clutter. Furthermore, the significance parameter selection is a fixed constant and visual interaction plays no role in its selection. In contrast, given the ubiquity of filaments in various density regimes, Felix allows for density ranged based probes into filaments, within clusters and voids. Furthermore, the visual interactive aspect allows for user engagement in parameter selection, which is crucial for the set of features identified. Another difference is that Felix uses simplification only for noise removal and not feature identification.

In this experiment, we demonstrate the consequences of not correlating the density characteristics for filament extraction. Specifically, we demonstrate that the filaments extracted using Felix are more spatially proximal to filament and cluster particles in the Voronoi Kinematic simulation. Furthermore, we show that tuning the significance parameter is not a sufficient mechanism to extract the desired filaments in this dataset. In the next experiment, we demonstrate the exploration of filaments within high-density cluster like regions and low-density void like regions. Such a delineation, coupled with the visual exploration process, is not possible using DisPerSE.

The recovery and failure rates for the Voronoi kinematic models A, B, and C are shown in Figure 4.9. The TP_d and FP_d of the Felix filaments are plotted in the top-left panel, the ones for DisPerSE can be found in the top right-hand panel (1σ significance threshold), bottom left-hand panel (3σ significance threshold) and bottom-right panel (5σ significance threshold).

Felix shows good recovery rates for all datasets, particularly around $d = 3h^{-1}\text{Mpc}$. For the least evolved configuration A, and for locations where structures are least



(a) (b) (c)

Figure 4.8 Filamentary structures extracted from datasets A, B and C of the Voronoi evolution time-series using DisPerSE with significance level of 5σ. Filaments are shown as orange tubes along with a volume rendering of the log-density field. The inset pictures show identified filaments that are within wall-like and void-like regions of the Voronoi kinematic datasets.

distinct, Felix still obtains moderately good recovery rates. The results for DisPerSE with simplification thresholds 1σ and 3σ are comparable: at short distances the true detection rate is slightly lower than that of Felix, while at larger distances it performs marginally better. The situation is slightly different in the case of DisPerSE with a 5σ simplification threshold. In the case of the more strongly evolved B and C datasets, DisPerSE and Felix have similar false detection rates FP_d, while the true detection rates TP_d of Felix are consistently higher.

For the lesser evolved datasets A and B, the false detection rates FP_d for DisPerSE quickly increase to rather large values. For simplification thresholds of 1σ and 3σ it even surpasses values of unity. This may indicate that in certain circumstances an automatic detection of filaments from the Morse-Smale complex runs the risk of over-determining the population of filaments, even after considerable simplification. While the problem is not so acute in the most evolved stage C, where the morphologies are well separated, direct simplification strategies may not always succeed in properly classifying all filament, wall and void regions in the more moderately evolved stages A and B.

Figure 4.8 shows the filaments in the three Voronoi models detected by DisPerSE, with a simplification threshold of 5σ. In comparison with the structures in Figure 4.7, the knot like structures present in clusters are absent. This leads to the cluster particles being far away from the filament end points, and thus the reduced TP_d rates of DisPerSE. In contrast, Felix's ranged query allows us to retain only the filaments in cluster like regions and those that connect these cluster like regions, leading to better TP_d rates. Also, many filaments found by DisPerSE are within the wall-like and void-like regions of the Voronoi Kinematic simulation. Some examples are highlighted using insets in Figure 4.8. Again these are filtered out by Felix's ranged query, which is not directly possible in DisPerSE. The inclusion of such structures in DisPerSE leads to its higher FP_d rates.

These findings suggest that a structure identification strategy based on a direct simplification procedure of the Morse-Smale complex should be applied with care to the density regimes being studied. Specifically, the superior classification rate pro-

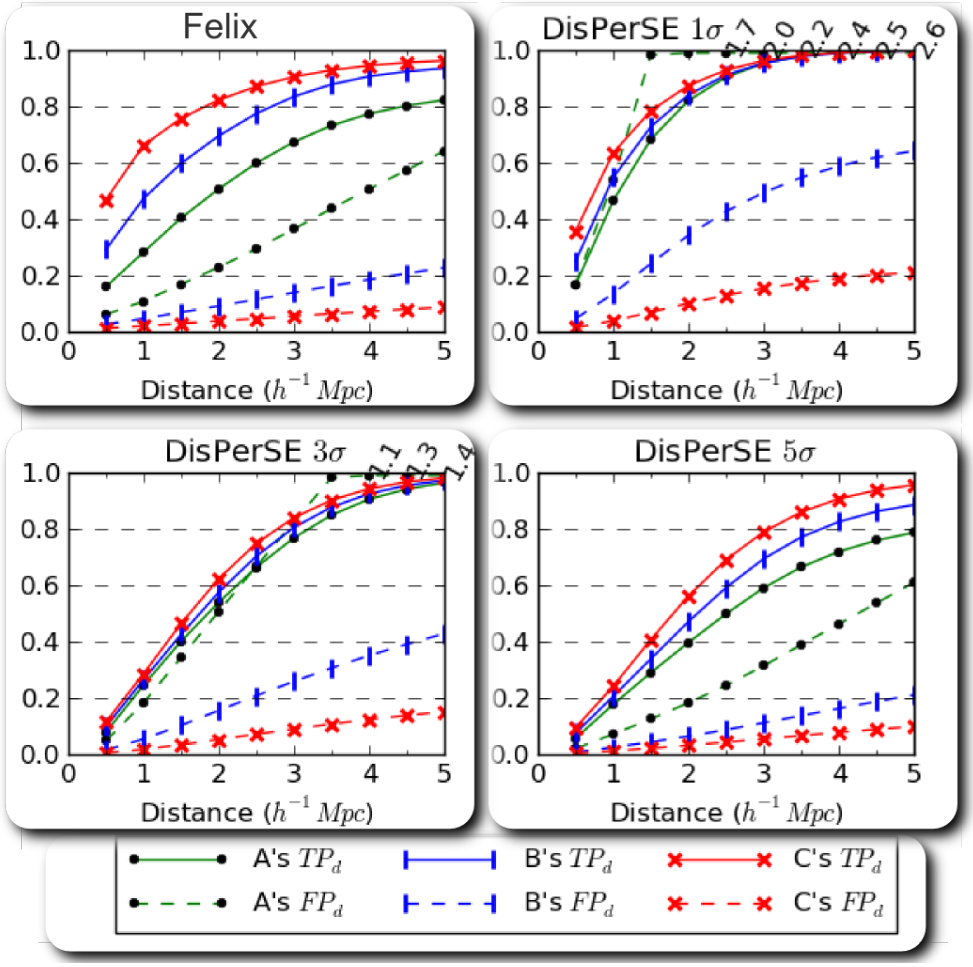


Figure 4.9 Classification recovery rates for Voronoi kinematic datasets A, B, and C. Top left: Felix; top right: DisPerSE, 1σ significance level; bottom left: DisPerSE, 3σ significance level, and (bottom right) DisPerSE with a 5σ significance level. False positive rates greater than 1.0 are clipped and respective values are shown.

files confirm that, using Felix, we can easily extract filaments that are spatially more proximal to the cluster and filament particles in the Voronoi Kinematic datasets. This issue is not easily addressed by the significance threshold of DisPerSE. In contrast, Felix provides an intutive density based handle to extract the desired features. Also, the possibility of having FP_d value larger than 1 in extreme situations is indicative of over-detection of filaments. This is potentially cumbersome for the analysis of genuine cosmological simulations and observational surveys. In more complex realistic circumstances, cosmic structure involves features over a wide range of densities and scales and structural morphologies that are not as well separated as in the simpler Voronoi models.

Dataset		TP_d	FP_d
A	$Felix(d = 3h^{-1}\text{Mpc})$	0.67	0.36
A	$DisPerSE(5\sigma, d = 3h^{-1}\text{Mpc})$	0.59	0.31
B	$Felix(d = 3h^{-1}\text{Mpc})$	0.84	0.13
B	$DisPerSE(5\sigma, d = 3h^{-1}\text{Mpc})$	0.69	0.11
	$Nexus/MMF$	0.85	0.13
C	$Felix(d = 3h^{-1}\text{Mpc})$	0.90	0.05
C	$DisPerSE(5\sigma, d = 3h^{-1}\text{Mpc})$	0.78	0.05
	$SpineWeb$	0.87	0.10

Table 4.2 Recovery rates of galaxies within $3h^{-1}$Mpc of structures extracted using Felix compared with Nexus/MMF, SpineWeb, and DisPerSE (5σ significance level).

Comparison B: Felix and SpineWeb

Spineweb (Aragón-Calvo et al. 2010) is also a technique based on Morse theory and exploiting the singularity structure of the density field. It exploits the fact that the watershed transform naturally outlines the basins around the local minima of the density field and directly defines the topological structure of the field, and as such forms a direct and practical tool for its analysis in terms of Morse theory. In this sense, SpineWeb is a direct development of the Watershed Void Finder WVF of (Platen et al. 2007), which introduced the definition of voids in terms of watershed basins.

As a preprocessing step, SpineWeb uses DTFE (Schaap & van de Weygaert 2000a; van de Weygaert & Schaap 2009a) to compute the density field on a regular grid from the given particle distribution. SpineWeb subsequently identifies the spine of the web as the regions that are excluded from the watershed basins. A given location in this spine is classified as belonging to a filament if it has three distinct void regions in its neighborhood. A region is classified as belonging to a wall if it has two and no more distinct void regions in its neighborhood.

For the comparison of the Felix and Spineweb (Aragón-Calvo et al. 2010), we use a Voronoi Evolution model realization that is comparable to the advanced state of dataset *C*. We use the test result reported in Aragón-Calvo et al. (2010) with respect to the model that has a similar percentage of particles in the four morphological features. In these model realizations, the clusters, filaments and walls have a Gaussian density profile with a scale of $R_g = 1h^{-1}$Mpc. The spine has an effective width $d = 2R_g$ as the identified structures are thickened by 1 voxel, with a size of $R_g = 1h^{-1}$Mpc.

For this configuration, Aragón-Calvo et al. (2010) report true and false detection rates of $TP_d = 0.87$ and $FP_d = 0.10$ (see Table 4.2). Felix attains the same recovery rate of $TP_d = 0.87$ at a smaller distance $d = 3h^{-1}\text{Mpc} = 1.5R_g$. For the same configuration, the failure rate parameter, $FP_d = 0.05$, is comparable to that reported for Spineweb. By comparison, at $d = 2R_g$, for Felix the recovery rates are $TP_d = 0.93$ and $FP_d = 0.07$.

In summary, these results appear to suggest that Felix performs as well as SpineWeb.

Comparison C: Felix and Nexus/MMF

The Nexus/MMF Multi-scale Morphology filter (Aragón-Calvo et al. 2007b; Cautun et al. 2013) identifies the morphology of regions on the basis of the ratio of the eigenvalues of the Hessian of the density field (MMF, Aragón-Calvo et al. (2007b)) or of fields that have a physical relation to the evolution and formation of the web-like structure in the cosmic mass distribution (Nexus,(Cautun et al. 2013)). The exact conditions for the classification as cluster, filament, wall and void can be found in (Aragón-Calvo et al. 2007b; Cautun et al. 2013).

A key aspect of Nexus/MMF is its explicit multi-scale nature. The density field is translated into a four-dimensional scale space map by convolving the density field with a Gaussian filter over a large range of different scales. The morphological identity of a give location is determined on the basis of a set well-defined morphological filters that compare the corresponding Hessian signature over a stack of convolved images in scale space.

As a result, we obtain a "scale-free" morphological assessment, in the sense that each region gets a unique classification tag based on a criterion that determines the most dominating morphology across the chosen range of scales.

Since Nexus/MMF concerns a formalism based on a scale space analysis, the parameters of detection do not have a direct correspondence with topology based techniques like DisPerSE, SpineWeb, and Felix.

We use dataset B for a comparison with Nexus/MMF. This dataset is similar to the least evolved dataset used by Aragón-Calvo et al. (2007b) in an evaluation of the MMF, the original density field based Nexus/MMF implementation. For similar values of the detection rate TP_d, both Felix and MMF have identical failure rates FP_d. This indicates that both procedures have a comparable detection behavior.

4.5.2 Filament Exploration

In this section we discuss the application of Felix to explore different classes of filaments from cosmological simulations. The ability to filter filaments on a combination of morphological and density properties is helpful in situations where we wish to focus on, for example, the properties of galaxies residing in filaments in low-density void regions or in the high-density outskirts of clusters. This ability of Felix to identify a specified population of intravoid filaments or cluster inflow channels provides us with a microscopic instrument that allows a detailed and systematic exploration of the fine structure in the hierarchy of cosmic structure. Felix is able to zoom in on such regions and delineate their detailed infrastructure.

As the criteria for the identification of filaments and other web-like features still differ substantially between the various available techniques, the visual interaction aspect of Felix is a major practical asset in obtaining a proper user-defined selection of filaments. To this end, we may also point out that available automatic detection techniques may produce significant spurious results, which may substantially influence the results of targeted studies as the one illustrated here. A telling example of this has been discussed in the previous section.

Figures 4.10a and 4.10c present volume renderings of a 3D region of the Cosmogrid dataset ranging from z-coordinates 69 to 105. The bounding box of the dataset

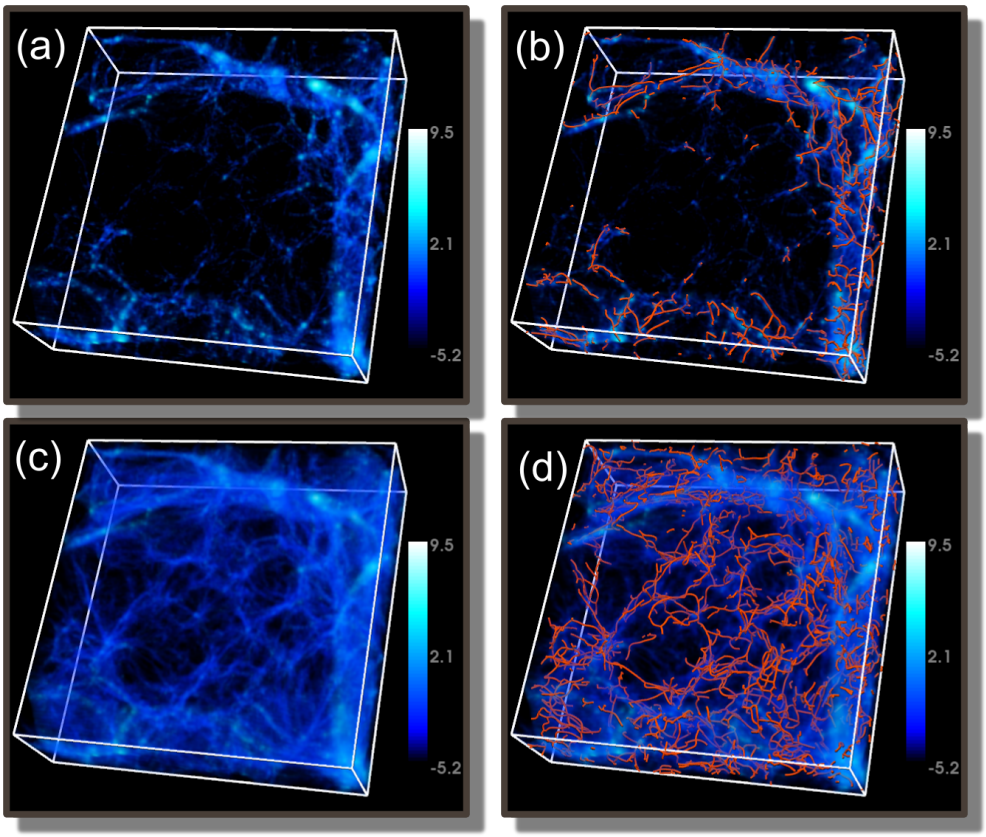


Figure 4.10 Exploring filaments in high-density cluster-like environments and low-density void like regions in the Cosmogrid dataset. The selected region, comprising of a 3D region of z-slices from 69 through 105, contains filamentary structures in both types of environments. Volume rendering of the density field with opacity adjusted to highlight filaments in (a) high density cluster-like regions, and (c) low density void-like regions. (b) Filaments within high density cluster-like regions extracted with parameters $[S_b, S_e] = [10^0, 10^{9.6}]$ and $[M_b, M_e] = [10^{2.3}, 10^{9.6}]$. (d) Filaments within low density void-like regions extracted with parameters $[S_b, S_e] = [10^{-2.5}, 10^{0.5}]$ and $[M_b, M_e] = [10^{0.5}, 10^{3.5}]$. For both sets of filaments, the value of T_s is set to 0.05.

is $128 \times 128 \times 128$. This region is selected as it contains a large void like region surrounded by a large number of high density regions. The transfer function opacities have been adjusted to highlight the filament like structures in cluster-like and void-like regions respectively. The hierarchical Morse-Smale complex computation and the filament selection are executed on the entire dataset. The resulting filaments and volume rendering are clipped to the above mentioned region of interest. The filament selection parameters of algorithm 1 are adjusted interactively, via a visualization step (see accompanying video). Each selection query takes approximately 1 second to process. The subsequent extraction of filament geometry depends on the number of selected 2-saddles. This takes approximately 4 seconds. Thus, the query framework may be used to interactively change parameters and visually correlate the set of extracted features with the underlying density distribution.

The filament selections obtained following the application of the interactive procedure are illustrated Figures 4.10b and 4.10d. Each shows filaments in a different environmental density regime. Figure 4.10b shows filaments that exist near and within the high density cluster like regions. These are the filaments that form the spine of the cosmic web. Figure 4.10d shows filaments within void-like regions. The combination of density criteria and interactive visualization enables us to zoom in on this system of intra-void filaments. They are the faint residuals of the smaller-scale filaments that constituted the spine of the cosmic web at earlier cosmic epochs, and as such represent a direct manifestation of the hierarchical buildup of cosmic structure. At the current epoch, the intra-void filaments appear to define a different pattern than the prominent filamentary bridges between clusters of galaxies. As a result of the tidal influence of surrounding large-scale mass concentrations they are conspicuously aligned along a direction correlated with the main axis of the embedding void.

4.5.3 Additional Cosmogrid experiment

In this section we discuss an additional application of Felix to explore three classes of filaments from the Cosmogrid simulation (Ishiyama et al. 2013). Figure 4.11a presents a snapshot of the particle distribution in a region of the simulation. Here, we use the 512^3 particle realization, which achieves a mass resolution of $8.21 \times 10^6 M_\odot$. The initial conditions are set up at $z = 65$ using the Zel'dovich approximation (Zel'dovich 1970). The region has been chosen such that it is characterized by a wide density range, multiple structural morphologies and low visual confusion. The particle distribution is used as input for computing a tetrahedral domain using a periodic triangulation and the density field is computed using the DTFE methodology (Schaap & van de Weygaert 2000a; van de Weygaert & Schaap 2009a).

The filament selections obtained following the application of the interactive procedure are illustrated in three panels. Each shows filaments in a different environmental density regime. Figure 4.11b shows filaments that exist near and within the high density cluster like regions, Figure 4.11c shows filaments within void-like regions and Figure 4.11d shows filaments that stretch all the way from cluster to void like regions. The latter are the filaments that form the spine of the cosmic web.

4.6 Volume rendering enhancement

In this section, we describe an application of our framework to suppress the opacity of regions spatially far removed from filamentary features of interest, which leads to a feature based volume rendering enhancement.

4.6.1 Volume Rendering

We use volume visualizations of the density field to aid selection of parameters for Algorithm 1. The geometry of the selected filaments using Algorithm 1 is superimposed upon a volume rendering of the density field. Based on the visualization of the extracted filaments and the density volume rendering, the parameters may be adjusted so that the structures correspond with the density volume rendering. Fig-

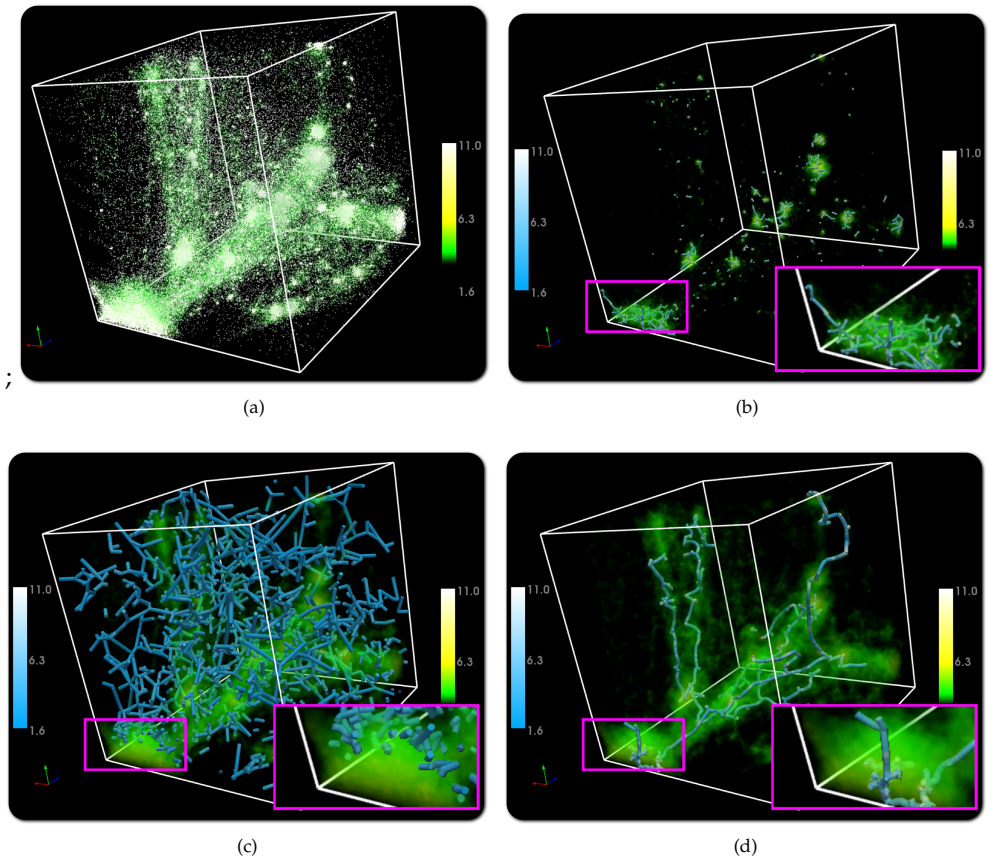


Figure 4.11 Three classes of filaments extracted from the Cosmogrid dataset. (a) Particle distribution shown along with a volume rendering of the DTFE density. (b) Filaments within cluster like regions extracted with parameters $[S_b, S_e] = [10^5, 10^8]$ and $[M_b, M_e] = [10^8, 10^{12}]$. The highlighted region shows the retention of intricate topological structures within a large cluster region. (c) Filaments within void like regions extracted with parameters $[S_b, S_e] = [10^0, 10^5]$ and $[M_b, M_e] = [10^0, 10^5]$. Shown in the inset is a cluster like region within which filamentary structures are filtered away as desired by the query to the framework. (d) Filaments that stretch from cluster like regions all the way down to void like regions using parameters $[S_b, S_e] = [10^3, 10^{11}]$ and $[M_b, M_e] = [10^9, 10^{11}]$. The highlighted region shows a single filament passing through the large cluster like region with intricate topological details filtered out by the framework.

ure 4.7 shows an example of the overlay of the volume visualization with the selected structures.

4.6.2 Methodology

Direct volume rendering of the cosmological density fields is often not effective for visualization because of the formation of clusters at multiple scales. Furthermore, these clusters are often spatially far removed from features of interest. Figure 4.12a shows the visual clutter caused in a cosmic web density field. Here, we discuss an application of Felix to suppress the opacity of regions spatially far removed from

filamentary features of interest, which leads to a feature based volume rendering enhancement. The input to our modified algorithm is the set of selected features, in addition to the density field and a transfer function that translates density values to color and opacity.

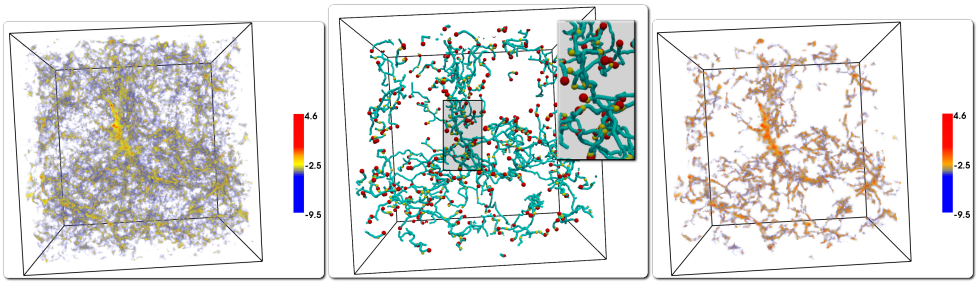
We modify the HAPT(Maximo et al. 2010) algorithm for enhancing the volume rendering of the density field defined on tetrahedral meshes. The HAPT algorithm projects tetrahedra in visibility order, while blending their raster pixels in a back-to-front order. The color and opacity of each raster pixel of a tetrahedron is given by the volume integral, which integrates color and opacity values along view rays as they enter and exit the tetrahedron. Moreland & Angel (2004) propose a linearized form of the integral for efficient evaluation and acceptable quality. The integral resolves into a linear expression with one non-linear term obtained from a pre-integrated table. A key advantage of this method is that the input required for each view ray is the color and opacity value at the end points of each view ray along with the length of the view ray through the tetrahedron. In the HAPT algorithm, the density at the vertices of the tetrahedron are first linearly interpolated to the endpoints of the view-ray tetrahedron intersection and then translated to color and opacity values using the transfer function.

In an analogous manner, we first compute the distance of the vertices from the selected set of features. This may be efficiently computed using a kd-tree (Bentley 1975) in $O(n \log n)$ time, where n is the number of vertices. Then, this distance is linearly interpolated to each end point of each view-ray tetrahedron intersection . The interpolated distance value at each end point of the view-ray tetrahedron is then used as an argument of a Gaussian to compute corresponding secondary opacity values. These secondary opacity values are multiplied by the opacity values obtained from the transfer function. The variance is used as a distance control parameter, where a lesser variance results in the suppression of the opacity of features that are spatially far removed from the selected set of features. The remainder of the HAPT algorithm is retained unchanged.

4.6.3 Subhaloes Dataset

For presenting the results of our volume enhancement routine, we use a large scale simulation in a box of $300h^{-1}$Mpc, that uses 512^3 particles described in details in (Bos et al. 2012; De Boni et al. 2011). The particles have a mass resolution of $4.43 \times 10^9 M_\odot$, and the initial conditions are set up at $z = 60$. For this experiment, we use the dark matter haloes instead of the particles. The halos consist of dark matter particles that clump together to form gravitationally bound, often virialized objects. The mean halo mass at $z = 0$ is $1.1 \times 10^{12} M_\odot$, with a halo mass range of $4.9 \times 10^{11} - 7.9 \times 10^{14} M_\odot$.

The dark matter haloes fairly accurately trace the patterns of the underlying mass distribution. We detect the haloes using one of the standard halo finding algorithms SUBFIND (Dolag et al. 2009). The first basic step involves the friends-of-friends method for finding halos. Particles are assigned to a group when they are within some linking length of any other particle in the group. This step is augmented by a sophisticated sequence of criteria for identifying gravitationally bound subclumps



(a) (b) (c)

Figure 4.12 Extracting filamentary features from the $300h^{-1}$Mpc ΛCDM dataset and enhancing the volume rendering. (a) Volume rendering of the density field. (b) Filaments selected by Algorithm SELECT2SADDLES with $[S_b, S_e] = [e^{-4.0}, \infty]$ and $[M_b, M_e] = [e^{1.0}, \infty]$ retains the filaments in high density regions. The intricate filamentary structure within a predominantly high density region is highlighted in the inset. (c) An enhancement produced by the volume rendering procedure applied to the selected filaments.

within the halo, ultimately producing a sample of halos and their subhalos in the simulated dark matter distribution.

4.6.4 Results

In this section, we illustrate the use of the volumetric enhancement routine in presenting a three-dimensional impression of the web-like network. The routine suppresses the depiction of components far removed from filaments of interest. One potential application of the enhancement algorithm is as a noise removal tool that is able to render the density field in the vicinity of structures of interest. This will help substantially towards elimination of the visual confusion and of possibly less interesting structures in the process.

For the experiment shown in figure 4.12, we use the the $300h^{-1}$Mpc ΛCDM dataset described above. This large-scale simulation provides a nice setting to illustrate the ability of the volume enhancement method. The dataset contains a multitude of structures spanning a wide range of density values that usually evoke a significant visual confusion. This may be clearly appreciated from the lefthand frame 4.12a. The subsequent filament selection and volume enhancement are shown in figure 4.12b. It provides a better appreciation of the spine of the filamentary network.

4.7 Conclusion

We have presented a Topology based Framework, named Felix, to probe filament structure in the large scale universe. The framework is particularly designed to probe filamentary structures in different density regimes, and optimally preserve structural detail in regimes of interest. While other cosmic structure analysis tools do not include a facility to select web-like features according to tailor-made aspects and characteristics, this is precisely the mission of the Felix procedure. We directed

Felix towards a case study of the the filamentary infrastructure and architecture of cosmic voids and demonstrated that it successfully extracts the network of tenuous filaments pervading their interior (van de Weygaert & van Kampen 1993; Sheth & van de Weygaert 2004; Weygaert 2007; Aragon-Calvo & Szalay 2013) .

In an accompanying study, we plan to exploit the Felix facility to study the physical characteristics of the extracted samples of intra-void filaments. This also involves their halo and subhalo population, their gas content, and the relation of these with the embedding voids and surrounding large-scale mass distribution. This will be of key importance towards understanding the formation and evolution of void galaxies (van de Weygaert et al. 2011; Kreckel et al. 2011, 2012) and specifically that of the issue of the missing dwarf galaxies (Peebles 2001). In addition, following the recognition that void architecture represents a potentially sensitive probe of dark energy and dark matter and a keen test of modified gravity theories (Ryden & Melott 1996; Schmidt et al. 2001; Park & Lee 2007a; Lavaux & Wandelt 2010; Bos et al. 2012; Lavaux & Wandelt 2012; Clampitt et al. 2013; Sutter et al. 2014b), the filament samples extracted by Felix will be subjected to a systematic study of their dependence on cosmological parameters.

As an immediate extension, we plan to use Felix with other scalar fields such as the tidal force field. Another possible direction is the visualization and analysis of the hierarchy of voids, walls, and filaments in cosmological datasets. Interactive visual exploration of these intricate structural networks remains a challenging and largely unexplored problem of major significance.

5

Discussions and Conclusions

Cosmology has entered the era of *big data* in the last decade. With the advent of new ground-based as well as space telescopes in commission or scheduled to be commision, the amount of data is in the order of petabytes and exabytes. On the one hand, the PLANCK mission and the PLANCK satellite are furnishing us with an unprecedented detailed measurement of the temperature ianisotropies in the Cosmic microwave background. On the other hand, ground based redshift surveys and telescopes like the Large Synoptic Survey Telescope (LSST) commisioned under the Dark Energy Survey (DES) project will trace billions or remote galaxies and provide multiple probes for the mysterious dark matter and dark energy. These surveys and telescopes are just a few to name in the plethora of missions planned to launch a coordinated attack on some of the recently discovered mysteries spewed at us. One of the most interesting among them is the realization that the Universe is expanding at an accelerating pace rather than slowing down. This has called for a renewed interest in the nature of dark energy and its properties. The recent influx of massive data in cosmology and related disciplines, calls for new methods of data analysis tailored to harness and extract the relevant information from the massive data in a systematic fashion.

Topological anaysis of cosmological datasets has a long history, and has been one of the principal pillars of investigation in the cosmological community for decades. This has mainly been achieved by describing and analyzing the features of the cosmic mass distrubution through Eluer characteristic, genus and Minkowsk functionals. There use also has been more heuristic in nature mainly aimed at discriminating between various models of cosmic mass distribution. While they have supplied a wealth of information on the nature of the topology and morphology of the cosmic mass distribution, there is a motivativation to introduce new topological methods for analysis, that tie into the salient features of the structure seen in the cosmos.

This thesis is motivated by recognizing the limitations of the existing methods in describing topology, to introduce new measures that are able to harness the topological information of the cosmic mass distribution in a greater detail. Topological data analysis, in particular persistence based analysis, of structural patterns has gained interest across various disciplines like medical imaging, cartography, agriculture etc., motivated by the similarity in the nature of the probelms approached. Morse theory and persistence based approach has also picked up interest in the cosmological community in the last few year, and has been applied to develop recipes for detecting and describing the structural patterns seen in the Cosmos. Noteworthy of metion among such methods are the SpineWeb formalism and the DiSperSe formalism (Aragón-Calvo et al. 2010; Sousbie 2011). These formalisms however have focussed solely within the scope of pattern recognition.

This thesis is a culmination of an interisciplinary collaboration beween the fields of cosmology, mathematics and computer science. On the front of cosmology, it has aimed to expand the scope of the analysis based on topological formalisms emanating from Morse theory, homology and persistence of the cosmic mass distribution. This has been through a comprehensive approach of integrating the quantitative topological analysis of models of cosmic mass distribution in terms of persistence and homology, with the more traditional approach of structure identification and detection. On the mathematical side, it has attempted to explore the general properties and characteristics of persistence homology and persistence diagrams. The properties of peristence diagrams are a topic of active research in the topo-mathematical community (Bubenik 2012). While the thesis has succeeded in establishing an empirical probabilistic view of persistence homology and diagrams through the introduction of the concepts of intensity and intensity maps, an analytical and theoretical framework has yet to be established. It is noteworthy to remember in this context is that a full analytical description of persistence may not be an easy challange, as has aptly been recognized in the wider mathematical community.

5.1 Future directions

5.1.1 Future directions

At the end of this thesis, there are a number of follow up investigations that come to mind. We list here a possible, but not exhaustive, set of directions that may be promising.

Statistics of persistence. We have established empirically that the ensemble average of persistence diagrams of stochastic processes are well-defined. An analytic probabilistic and statistical description of persistence will be a natural and important extension of the framework.

Persistence (Homology) of the large scale Universe. In this thesis, we presented a topological characterization of Gaussian random fields, with a view to understand the nature of fluctuations in the primordial Universe. Subsequently, we analyze the Voronoi models that mimic aspects of the matter distribution in the large scale Uni-

verse. A logical and important extension of this would be to investigate the topology of the genuine distributions of dark matter and halos in LCDM simulations.

Persistence characterization of the anisotropies in CMB. A persistence based hierarchical characterization of the anisotropies in the Cosmic Microwave Background, as provided by the latest measurements from the PLANCK satellite provides an interesting challenge Persistence has the potential to shed light into the hierarchical nature of primordial fluctuations. It may also be interesting to investigate the CMB with a view to isolate signatures of primordial non-Gaussianities. Using existing methodologies, the PLANCK team reports the absence of statistically significant non-Gaussian signals in the CMB maps. The hierarchical description of persistence has the potential to probe deeper in this direction.

Filament catalogues of the large scale Universe. Having tested the robustness of the filament finding software on test models and simulations, an important aspect to follow up is to produce filamentary catalogues from simulations as well as observation. These catalogs will have an option of querying for the properties of the associated galaxies they host, in view of investigating the formation and evolution of galaxies vis-a-vis the large scale environment they reside in.

Hierarchical characterization of Cosmic Voids. 2-dimensional persistence diagrams, that capture the formation and evolution of topological voids, can potentially provide a powerful tool for the characterization of the hierarchical aspects of cosmic voids. This is especially interesting, in view of the fact that topology, perhaps, is the only method that defines a void uniquely and unambiguously, without a choice of any free parameter. It has theoretically been suspected for long, and confirmed recently observationally that the cosmic voids form and evolve hierarchically. Studies have revealed cosmic voids are highly sensitive to dark energy and may contain information about modified gravity.

Characterization of the epoch of reionization. Due to its hierarchical nature, persistence is naturally tailored towards the study of reionization and the evolving network of ionization bubbles in the Epoch of reionization, marking the onset of formation of stars and galaxies. 0-dimensional persistence diagrams are optimally suited for capturing the evolving network of ionization bubbles, which are essentially isolated objects undergoing mergers and possibly more complex procedures. Additionally, studying the characteristics of the 1-dimensional diagrams will reveal details about the percolation properties of the distribution at that epoch. The 2-dimensional diagrams have a potential to reveal the evolving hierarchical network of voids in their infancy.

Appendices

A

Topology

A.1 Minkowski functionals

Suppose we have a solid body, $\mathbb{M}$, whose boundary is a smoothly embedded surface in $\mathbb{R}^3$. This surface may be a sphere or have holes, like the torus, and it may consist of one or several connected components, each with its own holes. Similarly, we do not require that $\mathbb{M}$ is connected. Write $\mathbb{M}^r$ for the set of points at distance r or less from $\mathbb{M}$. For small values of r, the boundary of $\mathbb{M}^r$ will be smoothly embedded in $\mathbb{R}^3$, but as r grows, it will develop singularities and self-intersections. Before this happens, the volume of $\mathbb{M}^r$ can be written as a degree-3 polynomial in r,

$$\mathrm{vol}\,\mathbb{M}^r \;=\; Q_0 + Q_1 r + Q_2 r^2 + Q_3 r^3\,. \tag{A.1}$$

The Q_i are known as the *Minkowski functionals* of $\mathbb{M}$, which are important concepts in integral geometry. For a d-dimensional manifold $\mathbb{M}$ there are $(d+1)$ Minkowski functionals. Minkowski functionals were first introduced as measures of the spatial cosmic mass distribution by Mecke et al. (1994) and have become an important measure of clustering of mass and galaxies (Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). In mathematics, they are closely related to concepts like the *Quermassintegrals*, *mixed volumes*, and *Killing-Lipschitz curvatures* $\mathcal{L}_i(\mathbb{M})$ in differential geometry. These names relate to different geometric interpretations of the Q_i, and in most cases involve a different ordering and normalization.

With respect to Killing-Lipschitz curvatures, we may observe that the expression above is the 3D version of *Weyl's tube formula* (Adler 1981). It is the general expression for the volume of the set of points that are at a distance $\leq r$ from an object $\mathbb{M}$ in d-dimensional space in terms of the Killing-Lipschitz curvatures (Adler & Taylor 2010;

Taylor & Adler 2009; Bobrowski & Borman 2012),

$$\text{vol}\,\mathbb{M}^r \;=\; \sum_0^d r^{d-i}\omega_{d-i}\mathcal{L}_i(\mathbb{M})\,. \tag{A.2}$$

In this expression ω_n is the volume of an n-dimensional unit sphere[1]. From this, we immediately find the identity between the Killing-Lipschitz curvatures $\mathcal{L}_{d-n}$ and the corresponding Minkowski functionals Q_n,

$$Q_n = \mathcal{L}_{d-n}/\omega_n\,. \tag{A.3}$$

In terms of their interpration in the three-dimensional context, following Equation A.1, we see that Q_0 is the volume of $\mathbb{M}$, Q_1 is the area of its boundary, Q_2 is the total mean curvature, and Q_3 is one third of the total Gaussian curvature of the boundary. These interpretations suggest that the Minkowski functionals are essentially geometric in nature, and they are, but there are strong connections to topological concepts as well. The key connection is established via the Euler characteristic, $\chi(S)$, of a surface S. We will discuss the latter notion shortly (see next subsection) but for now we just mention that the Euler characteristic – traditionally denoted as χ – is equal to 2 minus twice the number of holes. For example, the sphere has $\chi = 2$ and the torus has $\chi = 0$. If the boundary of $\mathbb{M}$ consists of k components with a total of h holes, then we have $\chi = 2(k-h)$. The connection between the topological characteristics of a manifold and its geometrical properties is stated by the famous Gauss-Bonnet theorem. For a connected surface S in $\mathbb{R}^3$, the Gauss-Bonnet theorem asserts that the total Gaussian curvature of a closed, is 2π times the Euler characteristic $\chi(S)$,

$$\chi(S) \;=\; \frac{1}{2\pi}\oint\left(\frac{1}{R_1R_2}\right)dS\,, \tag{A.4}$$

where R_1 and R_2 are the principal radii of curvature at each point of the surface. For the situation sketched above, a boundary of manifold $\mathbb{M}$ consisting of k components with a total of h holes, it tells that the total Gaussian curvature will be equal to $4\pi(k-h)$. Finally, $Q_3 = \frac{4\pi}{3}(k-h)$, i.e.

$$\begin{aligned} Q_3 \;&=\; \frac{2\pi}{3}\,\chi(S) \\ \mathcal{L}_0 \;&=\; \frac{1}{2}\,\chi(S)\,. \end{aligned} \tag{A.5}$$

For example, the Gaussian curvature of a sphere with radius r is $1/r^2$ at every point. Multiplying with the area, which is $4\pi r^2$, we get the total Gaussian curvature equal to 4π, which is independent of the radius. This agrees with $\chi = 4\pi(k-h)$ given above since $k-h=1$ in this case.

Important for our purpose is the observation that Minkowski functionals can be

[1] for the 3D situation, the relevant values of ω_k are $\omega_0 = 1$, $\omega_1 = 2$, $\omega_2 = \pi$ and $\omega_3 = \frac{4}{3}\pi$

expressed in terms of Euler integrals. The *Crofton intersection formula* of integral geometry (Crofton 1868) encapsulates a very profound statement on the length of curves, area of surfaces and a plethora of interesting geometric properties in terms of an integral over lower-dimensional intersecting hyperplanes. In a sense, it is the generalization of the famous Buffon's Needle problem (Ramaley 1969). The specific version of Crofton's formula pertaining to integrals over the Euler characteristic is known as *Hadwiger's Formula* (Hadwiger 1957; Adler & Taylor 2010). To evaluate the k-th Minkowski functional of a d-dimensional manifold $\mathbb{M}$, one has to consider the Euler characteristic of the intersection of k-dimensional hyperplanes S_k with $\mathbb{M}$, $\chi(S_k \cap M)$. The value of the Minkowski functional $Q_k(\mathbb{M})$ is equal to the integral of the Euler characteristic $\chi(S_k \cap M)$ over the space $\mathcal{E}_k^d$ of all conceivable hyperplanes S_k (Schmalzing & Buchert 1997),

$$Q_k(\mathbb{M}) \;=\; \frac{\omega_d}{\omega_{d-k}\omega_k} \int_{\mathcal{E}_k^d} \mathrm{d}\mu_k(S_k)\, \chi(S_k \cap M)\,, \tag{A.6}$$

with the normalization constants ω_j are the volumes of j-dimensional unit spheres.

A.2 Euler characteristic and genus

Suppose now that we have the boundary of $\mathbb{M}$ triangulated, using v vertices, e edges, and t triangles. Named after Leonhard Euler (Euler 1758), the *Euler characteristic* of the surface is the alternating sum of the number of simplices:

$$\chi = v - e + t\,. \tag{A.7}$$

It does not depend on the triangulation, only on the surface. For example, we can triangulate the sphere with 4 vertices, 6 edges, and 4 triangles, like the boundary of the tetrahedron, which gives $\chi = 4 - 6 + 4 = 2$. Alternatively, we may triangulate it with 6 vertices, 12 edges, and 8 triangles, like the boundary of the octahedron, which again gives $\chi = 6 - 12 + 8 = 2$.

As mentioned above, the Euler characteristic of a connected, closed surface with $h \geq 0$ holes is $\chi = 2 - 2h$. To make this more concrete, we formalize the number of holes of a closed, connected surface to its *genus*, denoted as $g = h$. It is defined as the maximum number of closed curves we can draw on the surface such that cutting along them leaves the surface in a single connected piece. For example, for a sphere we have $g = 0$, and for a torus we have $g = 1$. If we now drop the assumption that the surface is connected, we get the Euler characteristic and the genus by taking the sum over all components. Since $\chi_i = 2 - 2g_i$ for the i-th component, we have

$$\chi = \sum_{i=1}^{k} \chi_i = \sum_{i=1}^{k} (2 - 2g_i) = 2k - 2g\,. \tag{A.8}$$

We see that a minimum amount of topological information is needed to translate between Euler characteristic and genus. This is different from what the cosmologists

have traditionally called the genus, which is defined as $\tilde{g} = -\frac{1}{2}\chi$. Relating the two notions, we get $g = k + \tilde{g}$. We will abandon both in this paper, $\tilde{g}$ because it is redundant, and g because it is limited to surfaces. Indeed, the Euler characteristic can also be defined for a 3-dimensional body, taking the alternating sum of the simplices used in a triangulation, while the genus has no satisfactory generalization beyond 2-dimensional surfaces.

The Gauss-Bonnet theorem (eq. A.4) and Crofton's formula (eqn. A.6) underline the key position of the Euler characteristic at the core of the topological and geometric characterization of manifolds. The Euler characteristic establishes profound and perhaps even surprising links between seemingly widely different areas of mathematics (Adler & Taylor 2010). While in simplicial topology Euler's polyhdron formula states that it is the alternating sum of the number of k-dimensional simplices of a simplicial complex (eq. A.7), its role in *algebraic topology* as the alternating sum of Betti numbers is expressed by the Euler-Poincaré formula (see eq. A.9 in the next subsection). Even more intricate is the connection that it establishes between these topological aspects and the singularity structure of a field, which is the realm of *differential topology*. In particular interesting is the relation established by Morse theory of the Euler characteristic being equal to the alternating sum of the number of different field singularities, ie. of maxima, minima and saddle points. Finally, its significance in *integral geometry* is elucidated via Crofton's formula (eq A.6), which establishes the fact that Minkowski functionals are integrals over the Euler characteristic.

A.3 Homology and Betti numbers

While the Euler characteristic can distinguish between connected, closed surfaces in $\mathbb{R}^3$, it has no discriminative power if applied to 3-manifolds, which is the most direct generalization of surfaces to the next higher dimension. Indeed, Poincaré duality implies $\chi = 0$ for all 3-manifolds. Fortunately, we can write the Euler characteristic as an alternating sum of more descriptive topological invariants named after Enrico Betti (Betti 1871). To introduce them, we find it convenient to generalize the space $\mathbb{M}$ by dropping most limitations, such as that it be embedded or even embeddable in $\mathbb{R}^3$. Letting the intrinsic dimension of $\mathbb{M}$ be d, we get $d+1$ possibly non-zero *Betti numbers*, which traditionally are denotes as $\beta_0, \beta_1, \ldots, \beta_d$. The relationship to the Euler characteristic is give by the Euler-Poincaré Formula:

$$\chi \;=\; \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{A.9}$$

This relation holds in great generality, requiring only a triangulation of the space, and even this limitation can sometimes be lifted. In this paper, we only consider subspaces of the 3-torus: $\mathbb{M} \subseteq \mathbb{X}$. For this case, only β_0, β_1, β_2, and β_3 are possibly non-zero, and we have $\beta_3 \neq 0$ only if $\mathbb{M} = \mathbb{X}$, in which case $\beta_3 = 1$. The first three Betti numbers have intuitive interpretations: β_0 is the number of *components*, β_1 is the number of *loops*, and β_2 is the number of *shells* in $\mathbb{M}$. Often, it is convenient to consider the complement of $\mathbb{M}$, which shows $\beta_0 - 1$ *gaps* between the components, β_1 *tunnels* going through the loops, and β_2 *voids* enclosed by the shells.

A formal definition of the Betti numbers requires the algebraic notion of a ho-

mology group. While a serious discussion of this topic is beyond the scope of this paper, we provide a simplified exposition and refer to texts in the algebraic topology literature for details (see e.g. Munkres 1984).

For simplicity, we assume a triangulated space and we use the coefficients 0 and 1 and addition, modulo 2. A *p-chain* is a formal sum of the p-simplices in the triangulation, which we may interpret as a subset of all p-simplices, namely those with coefficients 1. The *sum* of two p-chains is again a p-chain. Interpreted as sets, the sum is the symmetric difference of the two sets. Note that each p-simplex has $p + 1$ (p-1)-simplices as faces. The *boundary* of the *p-chain* is then the sum of the boundaries of all p-simplices in the chain. Equivalently, it is the set of (p-1)-simplices that belong to an odd number of p-simplices in the chain. We call the p-chain a *p-cycle* if it is the boundary of a (p+1)-chain. Importantly, every p-boundary is a p-cycle. The reason is simply that the boundaries of the (p-1)-simplices in the boundary of a p-simplex contain all (p-2)-simplices twice, meaning the boundary of the boundary is necessarily empty. To get homology, we still need to form *classes*, which we do by not distinguishing between two p-cycles that together form the boundary of a $(p + 1)$-chain.

To get the group structure, we add p-cycles by taking their symmetric difference or, equivalently, by adding simplices modulo 2. Homology classes can now be added simply by adding representative p-cycles and taking the class that contains the sum. The collection of classes together with this group structure is the *p-th homology group*, which is traditionally denoted as H_p. Finally, the *p-th Betti number* is the rank of this group, and since we use modulo 2 arithmetic to add, this rank is the binary logarithmic of the order: $\beta_p = \log_2 |H_p|$. We note that modulo 2 arithmetic has multiplicative inverses and therefore forms what in algebra is called a *field*. For example, arithmetic with integers is not a field. Whenever we use a field to construct homology groups, we get vector spaces. In particular, the groups H_p defined above are vector spaces, and the β_p are their dimensions, as defined in standard linear algebra.

In our study, we forward Betti numbers for the characterization of the topological aspects of the cosmic mass distribution. In this context, we should also appreciate the significance of the observation that Minkowski functionals may be written as Euler integrals, expressed by Hadwiger's formula (eqn. A.6). Wintraecken (2012) recently demonstrated that Betti numbers, as opposed to Minkowski functionals, cannot be expressed in terms of integrals over the Euler characteristic. The importance of this finding for our purpose is that Betti numbers contain topological information which is different and complementary to that contained in Minkowski functionals and genus, and that an analysis of their characteristics will shed new light on the connectivity of the different morphological elements of the cosmic web.

A.4 Morse theory

In Morse theory, we consider a compact manifold $\mathbb{X}^\nu$,and a generic smooth function on this manifold. In the context of this paper, the manifold is the 3-torus and the function is a density distribution, $\varrho : \mathbb{X} \to \mathbb{R}$. Assuming ϱ is smooth, we can take derivatives, and we call a point $x \in \mathbb{X}$ *critical* if all partial derivatives vanish. Corre-

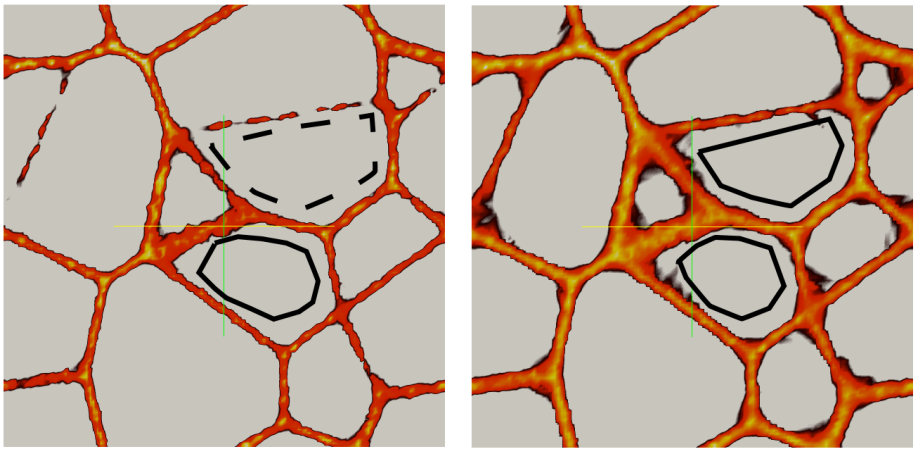


Figure A.1 Density rendering of the superlevel set of a 2-dimensional cross section of the voronoi wall models, illustrating chains and cycles. We focus our attention on the structures traced by the black lines. For high superlevel set values, in the left panel, the structure traced by the broken D-shape does not form a loop. The multiple broken segments are all *chains*. For lower superlevel set, the structures thicken and the individula segments merge together to form a loop, a 1-dimensional *cycle*.

spondingly, $\varrho(x)$ is a *critical value* of the function. All points of $\mathbb{X}$ that are not critical are *regular points*, and all values in $\mathbb{R}$ that are not the function value of critical points are *regular values*. Finally, we call ϱ *generic* if all critical points are non-degenerate in the sense that they have invertible Hessians. In this case, critical points are isolated from each other, and since $\mathbb{X}$ is compact, we have only finitely many critical points and therefore only finitely many critical values. The *index* of a non-degenerate critical point is the number of negative eigenvalues of the Hessian. Since $\mathbb{X}$ is 3-dimensional, we have 3-by-3 Hessians and therefore only four possibilities for the index. A *minimum* of ϱ has index 0, a *maximum* has index 3, and there are two types of *saddles*, with index 1 and 2.

The significance of the critical points and their indices becomes apparent when we look at the sequence of growing superlevel sets: $\mathbb{X}^\nu = \varrho^{-1}[\nu, \infty)$, for $0 \leq \nu < \infty$. If $\nu > \mu$ are regular values for which $[\mu, \nu]$ contains no critical value then $\mathbb{X}^\nu$ and $\mathbb{X}^\mu$ are topologically the same, the second obtained from the first by diffeomorphic thickening all around. If $[\mu, \nu]$ contains the critical value of exactly one critical point, x, then the difference between the two superlevel sets depends only on the index of x. If x has index 3 then $\mathbb{X}^\mu$ has one more component than $\mathbb{X}^\nu$, and that component is a topological ball. If x has index 2, then $\mathbb{X}^\mu$ can be obtained from $\mathbb{X}^\nu$ by attaching an arc at its two endpoints and thickening all around. This extra arc can have one of two effects on the homology of the superlevel set. If its endpoints belong to different components of $\mathbb{X}^\nu$, then $\mathbb{X}^\mu$ has one less component, while otherwise $\mathbb{X}^\mu$ has one

more loop. If x has index 1, then $\mathbb{X}^\mu$ can be obtained from $\mathbb{X}^\nu$ by attaching a disk, which has again one of two effects on the homology groups. Finally, if x has index 0 then $\mathbb{X}^\mu$ is obtained by attaching a ball. In all cases but one, this ball fills a void, the exception being the last ball that is attached when we pass the global minimum of ϱ. At this time, the superlevel set is completed to $\mathbb{X}^\mu = \mathbb{X}$.

A.5 Persistence homology

In Morse theory, we learned that each critical point either increases the rank of a homology group by one, or it decreases the rank of another group by one. Equivalently, it gives birth to a generator of one group or death to a generator of another group. Our goal is to pair up births with deaths such that we can talk about the subsequence in the filtration over which a homology class exists. This is precisely what persistent homology accomplishes. The hierarchical definition of topology that emerges due to taking the path of filtration is ideally suited to describe the topology of the mass distribution in the Universe, on account of it being hierarchical in nature as well.

To describe how this is achieved, we map each superlevel set in the filtration to the direct sum of its homology groups. With this construction, we capture the homology classes of all dimensions as once, and we simplify the notation by making it unnecessary to write the dimension in the subscript. Recall that between two consecutive critical values, the homology of the superlevel sets is constant. It therefore suffices to pick one regular value within each such interval. Writing $r_0 > r_1 > \ldots > r_n$ for these regular values and H_i for the direct sum of all homology groups of the superlevel set $\mathbb{X}^{r_i} = \varrho^{-1}[r_i, \infty)$,we get a sequence of homology groups:

$$H_0 \to H_1 \to \ldots \to H_n,$$

where $H_0 = 0$ and H_n is the homology group of $\mathbb{X}$. The arrows represent homomorphisms induced by the inclusion between superlevel sets. Assuming coefficients in a field, as before, we have a sequence of vector spaces with linear maps between them. These maps connect the groups by telling us where to find the cycles of a homology group within later homology groups. Sometimes, there are new cycles that cannot be found as images of incoming maps, and sometimes classes merge to form larger classes, which happens when we get chains that further wash out the difference between cycles.

We are now more specific about these connections. Letting γ be a class in H_i, we say γ is *born* at H_i and *dies entering* H_j if

- γ is not in the image of H_{i-1} in H_i;
- the image of γ is not in the image of H_{i-1} in H_{j-1}, but it is in the image of H_{i-1} in H_j.

Letting $r_{i-1} > \nu_i > r_i$ and $r_{j-1} > \nu_j > r_j$ be the critical values in the relevant intervals, we represent γ by (ν_i, ν_j), which we call a *birth-death pair*. Furthermore, we call $\operatorname{pers}(\gamma)$ the *persistence* of γ, but also of its birth-death pair.

To avoid any misunderstanding, we note that there is an entire coset of homology classes that are born and die together with γ, and all these classes are represented

by the same birth-death pair. Calling the image of H_i in H_{j-1} a *persistent homology group*, we note that its rank is equal to the number of birth-death pairs (ν_b, ν_d) that satisfy $\nu_b \geq \nu_i > \nu_j \geq \nu_d$. They represent the classes that are born at or before H_i and that die entering H_j or later.

B

Computation

The geometric and topological concepts outlined in Appendix A have all matured to a stage at which we have fast software to run on simulated and observed data. In this section, we describe the principles of these algorithms, and we provide sufficient information for the reader to understand the connection between the mathematics, the data, and the computed results.

The computational framework of our study involves three major components. One concerns the definition and calculation of the density field on which we apply the field's filtration. A directly related issue is the representation of the density field in the homology calculation, ie. whether we retain its representation by density estimates at the original sampling points or whether we evaluate it on the basis of a density image on a regular grid. The procedure for constructing the filtration is different in each case, and we detail them in Section B.1 and Section B.2. The algorithm used to implement the actual homology computation is the core of our study, and is same for both the filtration defined on particles or a filtration defined on image data. The third major aspect of our study concerns the representation of the results of the homology computation. The principal analytical tools of our study consist of *intensity maps* and *Betti numbers*, which form the visual representation and summary of the persistent homology of the analyzed data samples.

B.1 Density filtration from point samples

We use DTFE (Schaap & van de Weygaert 2001; van de Weygaert & Schaap 2009b; Cautun & van de Weygaert 2011) to construct a piecewise linear scalar-valued density field from a particle distribution. The DTFE formalism, whose details are outlined in appendix F.0.6, involves the computation of the Delaunay triangulation of

Model	# particles	# simplices	Tri. (s)	Pers. (s)
Poisson	500,000	14,532,164	10.15	6414.16
Cluster	262,144	7,491,308	81.48	12.58
Filament	262,144	7,346,712	77.76	402.36
Wall	262,144	7,345,520	5.26	555.46
Voronoi Kinematic Stage 3	262,144	7,409,364	5.93	125.33
Soneira-Peebles $\zeta = 9.0$	531,441	14,300,836	162.42	168.15

Table B.1 Parameters of computation for the various models described in this paper. All computations are performed on an Intel(R) Xeon(R) CPU @ 2.00GHz. Columns 1 & 2 present the models described in the later sections, and the number of particles used for the computation. Column 3 gives the total number of simplices of the Delaunay triangulation. Columns 4 & 5 give the time required to compute the triangulation and persistence respectively, in seconds.

the particles in $\mathbb{X}$, the determination of tessellation based density estimates and the subsequent piecewise linear interpolation of the density values at the Delaunay vertices, ie. the sample points, to the higher dimensional simplices, yielding a field $\varrho : \mathbb{X} \to \mathbb{R}$.

For the calculation of the Delaunay tessellation, we use software in the CGAL library. We use the 3-torus option of CGAL, which is the the periodic form of the original data set in a cubic box, which is accomplished by identifying and glueing opposite faces of the box.

Table B.1 presents the noteworthy parameters of computations for a single realization of the different models used in the results section of this paper. Naming the models in Column 1, we see the number of particles and simplices in the Delaunay triangulation in Columns 2 and 3 (also see Okabe et al. 2000; van de Weygaert 1994), and the number of seconds needed to compute the Delaunay triangulation and the persisten pairs in Columns 4 and 5. Apparently, the number of particles is not strongly correlated with the time it takes to construct the Delaunay triangulation. Indeed, the algorithm is also sensitive to other parameters – such as the number of simplices in the final triangulation or ever constructed and destroyed during the runtime of the algorithm – that depend on how the particles are distributed in space.

In a second step, we compute the DTFE density value for each vertex, u, of the Delaunay triangulation. The DTFE density value at the vertices is the inverse of the volume of its *star*. The star consists of all simplices that contain u as a vertex (see Figure B.1 for an illustration) , and we assign one over this volume as the density value to u. Finally, we use piece-wise linear interpolation to define $\varrho : \mathbb{X} \to \mathbb{R}$.

Evidently, we should ask ourselves in how far the results of the homology analysis are dependent on the density estimator used. In Section 2.9 we present a com-

parison between the results of our homology analysis obtained using the DTFE density estimate and that using the spatially adaptive SPH density estimates (see Appendix F.0.7), demonstrating that the results based on DTFE and on SPH are largely consistent.

B.2 Density filtration of fields on a regular grid

A first step towards computing persistence of a field ρ sampled on a regular cubical grid is the construction of a triangulation on the sample voxels. The components of a triangulation - vertices, edges, faces and tetrahedra - define a simplicial complex whose topological characteristics are equivalent to that of the sampled field. It is not possible to construct a unique triangulation K from a regular cubical grid of sample voxels. This is because a cubical grid suffers from degeneracies caused by corners common to eight and not four voxels, and the edges shared by four and not three faces.

Bendich et al. (2010) provide an algorithm for constructing triangulations of images represented on regular voxel grids. The algorithm solves the degeneracy arising from the regular grid by slightly perturbing the grid cells leading to a deformed grid where the corners are shared by four voxels, and the edges by three voxels. This transformation defines the elements of the dual triangulation uniquely – the vertices of this triangulation are defined by voxel centers, the edges are defined by the centers of the voxels which share a common face, the triangles by the centers of the voxels which share a common edge, and the tetrahedra by the centers of the voxels which share a common corner.

In a second step of the algorithm, the field values at the vertices in the triangulation are used to interpolate the values on the higher dimensional simplices, much akin to that used in the DTFE formalism developed by Schaap & van de Weygaert (2000b) (also see (Bernardeau & van de Weygaert 1996; van de Weygaert & Schaap 2009b)). This results in a continuous linearly extrapolated simplicial field - ie. a field defined on the edges, faces and tetrahedra of the resulting simplicial complex - that preserves the topology of the original density field (see (Pranav et al. 2013)). Of crucial importance is the fact that the choice of interpolation - linear, or constant - has no effect on topology. In this paper we we use a piece wise constant interpolation: $\rho(\sigma) = max[\rho(\tau)|\tau \subset \text{vertex of} \sigma]$.

B.3 Field sampling

To further process the density field towards its topological analysis, we may follow a range of field sampling strategies. The most suggestive option is to take the raw DTFE field, including all details of the discretely sampled field. In the context of our study, we mostly follow this *Raw DTFE sampling* strategy. The particular nature of the discretely sampled density field involves a complication. Because the number density of the sample points represents a measure of the value of the density field itself, the DTFE density field has a much higher spatial resolution in high density regions than in low density regions. This might be a source of a strong bias in the re-

trieved topological information, given that most of this will focus on the topological structure of the high-density regions.

To alleviate a density bias in the topological analysis, one may invoke a range of strategies. One option is to sample the density field on a regular grid. In other words, to create an image of the DTFE density field reconstruction. Details of the image construction are described in appendix F.0.6. It has the advantage of representing a uniformly sampled density field, with a uniform spatial resolution dictated by the voxel size of the image. The homology analysis of such a gridbased image involves a few extra complications, the details of which are most extensively discussed in the follow-up study analyzing the homology of Gaussian random fields (see Chapter 3). In appendix 2.9 we have compared the results of the homology analysis involving this *DTFE image sampling* strategy with those obtained following the *Raw DTFE sampling* strategy.

Another strategy to moderate the bias towards high-density regions is to use the singularity structure of the piecewise linear density field, and use the persistence of singularity pairs to remove insignificant topological features. This natural feature-based smoothing of the density field has been described extensively by Edelsbrunner et al. (2003), and has been applied in studies of cosmic structures in Chapter 4.

B.4 Topology through critical points and filtrations.

As mentioned in the paragraph on Morse theory, the superlevel set does not change topology as long as ν does not pass a critical value of the function, and this is also true for piecewise linear functions, except that we need to adjust the concept of critical point. Here we do the obvious, looking at how ϱ varies in the *link*, of a vertex. The link consists of all faces of simplices in the star, that do not themselves belong to the star (Edelsbrunner & Harer 2010, Chapter VI). Indeed, the topology can change only when ν passes the value of a vertex, so it suffices to consider only one (regular) value between any two contiguous vertex values. To describe this, we let n be the number of vertices in the triangulation, and we assume $\nu_i = \varrho(u_i) < \nu_{i+1} = \varrho(u_{i+1})$ for $1 \leq i < n$.[1] We thus consider superlevel sets at the regular values in the sequence

$$r_0 > \nu_1 > r_1 > \nu_2 > \ldots > \nu_n > r_n.$$

Constructing these superlevel sets and computing their homology individually would be impractical for the data-sets we study in this paper. Fortunately, there are short-cuts we can take that speed up the computations while having no effect on the computed results. The first short-cut is based on the observation that $\mathbb{X}^\nu$ has the same homotopy type as the subcomplex K^ν of the triangulation K of $\mathbb{X}$ that consists of all vertices with $\varrho(u_i) \geq \nu$ and all simplices connecting them. There is a convenient alternative description of K^ν. Define the *upper star* of a vertex u as the collection of simplices in the star for which u is the vertex with smallest density value (see fig. B.1 for the upper star of a regular vertex, a 1-saddle, a 2-saddle an a maximum). Then K^ν

[1] It is unlikely that the estimated density values at two vertices are the same, and if they are, we can pretend they are different, eg. by simulating a tiny perturbation that agrees with the ordering of the vertices by index; see eg. (Edelsbrunner 2001, Section I.4).

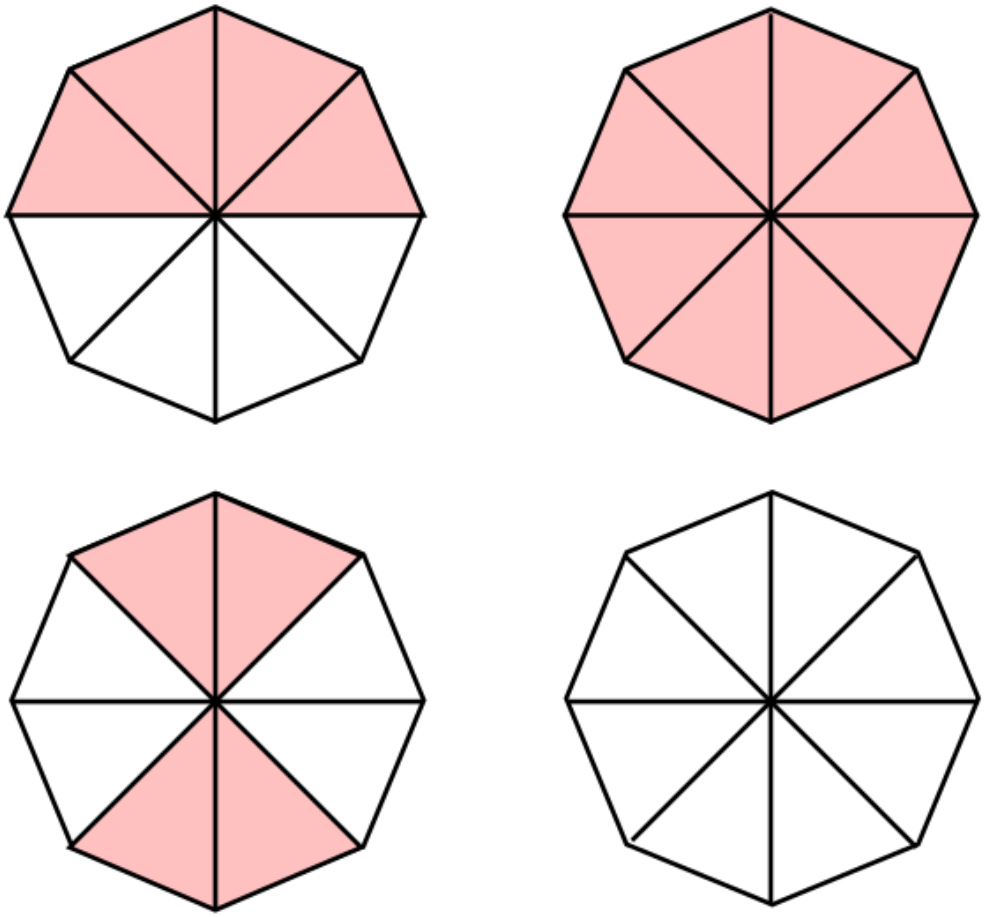


Figure B.1 Figure illustrating the upper star of a regular vertex, minimum, saddle and maximum, respectively in top-left, top-right, bottom-left, bottom-right panels. The star of a vertex is consists of all the simplices incident to it. The shaded simplices in pink have a function value higher than the vertex.

is the union of the upper stars of all vertices with $\varrho(u_i) \geq \nu$. This description is computational convenient because it tells us that $K^{r_{i+1}}$ can be obtained from K^{r_i} simply by adding the simplices in the upper star of u_{i+1}. We say the superlevel sets can be computed *incrementally*, and we will be careful to follow this paradigm in every step of our computational pipeline. This incremental construction of the superlevel sets is equivalent to constructing the upper-star filtration, which is an essential pre-cursor to computing persistence homology which we describe next.

B.5 Persistence homology

Next, we sketch the algorithm that computes the persistent homology of the sequence of superlevel sets. We begin with a linear ordering of the simplices in K that contains all K^v as prefixes. To describe it, let $u_i = \sigma_{j_i}, \sigma_{j_i+1}, \ldots, \sigma_{j_{i+1}-1}$ be the simplices in the upper star of u_i, sorted in increasing order of dimension. Setting $j_1 = 1$ and $m = j_{n+1} - 1$, this linear ordering of the simplices is $\sigma_1, \sigma_2, \ldots, \sigma_m$. It has the property that each simplex is preceded by its faces, which implies that every prefix, $K_j = \{\sigma_1, \sigma_2, \ldots, \sigma_j\}$, is a simplicial complex. We require this property so that every step of our incremental algorithm is well defined. It should be clear that $K^{v_i} = K_j$ for $j = j_{i+1} - 1$.

Algorithm 2 MATRIX REDUCTION

```
1: R = Δ
2: for j = 1 to m do
3:    while there exists j_0 < j with low(j_0) = low(j) do
4:       add colum j_0 to column j
5:    end while
6: end for
```

B.5.1 Boundary matrix and its reduction

The persistence algorithm is easiest to describe as a matrix reduction algorithm, with the input matrix being the ordered boundary matrix of K.[2] Specifically, this is the m-by-m matrix Δ whose rows and columns correspond to the simplices in the mentioned linear ordering. Specifically, the j-th column records the boundary of σ_j, namely $\Delta_{i,j} = 1$, if σ_i is a face of σ_j and the dimension of σ_i is one less than that of σ_j, and $\Delta_{i,j} = 0$, otherwise. Symmetrically, the i-th row records the star of σ_i. The persistence algorithm transforms Δ into *reduced form*, in which every row contains the lowest non-zero entry of at most one column. Making sure that we do not permute rows, and we add columns strictly from left to right, the lowest non-zero entries in the reduced matrix correspond to the birth-death pairs of the density field – precisely the information we are after. To describe the transformation, we write $\text{low}(j) = i$ if i is the maximum row index of a non-zero entry in column j, and we set $\text{low}(j) = 0$ if the entire column is 0. Algorithm 2 presents the algorithm for such a reduction. Section 2.3.3 illustrates these concepts and steps through an example.

The search for the fastest algorithm to reduce an ordered boundary matrix is an interesting question of active research in the field of computational topology. Most known algorithms use row and column operations, like in Gaussian elimination, which takes time proportional to m^3 in the worst case. A fortunate but largely not understood phenomenon is the empirical observation that some of these algorithms

[2] We hasten to mention that storing this matrix explicitly is too costly for our purposes. Instead, we use the triangulation as a sparse matrix representation, and we implement all steps of the matrix reduction algorithm accordingly. However, for the purpose of explaining the algorithm, we maintain the illusion of an explicit representation of the matrix.

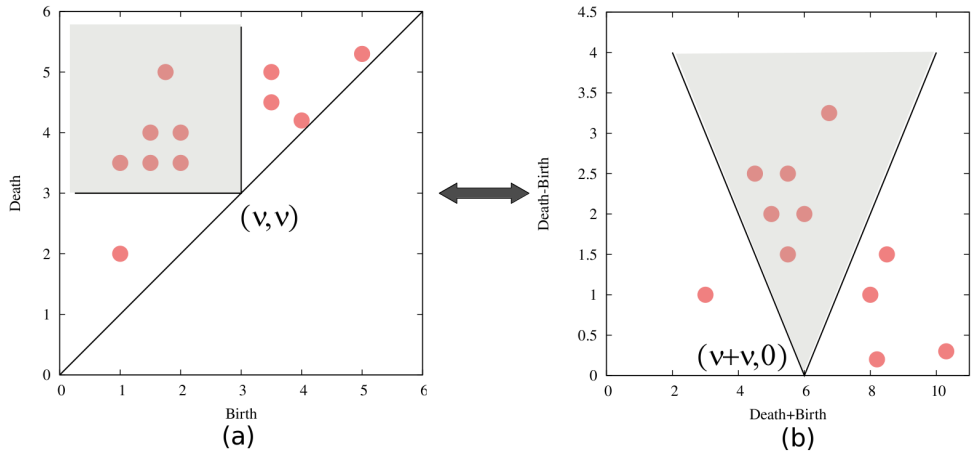


Figure B.2 Figure illustrating the transition from the *birth-death* to the *mean age-persistence* plane. If the coordinates of a point in panel (a) are (b,d), the coordinates in panel (b) are (d+b,d-b). The Betti numbers can be read off from the persistence diagrams. The contribution to the Betti numbers for a level set ν comes from all the persistent dots that are born before ν and die after ν – in other words, the shaded region in panel (a) anchored at (ν, ν). The shaded region transforms in panel (b) to a V-shaped region anchored at $(\nu + \nu, 0)$. The arms of the V have slope -1 and 1 respectively.

are significantly faster than cubic time for most practical input data. This is lucky but also necessary since we could otherwise not compute the results we present in this paper. The time to compute the persistence pairs for different models is displayed in Column 5 of Table B.1.

B.6 Persistence Diagrams

Given the reduced boundary matrix, we generate the birth-death pairs of ϱ from the lowest non-zero entries in the columns. Specifically, for every non-zero $i' = \mathrm{low}(j')$, the addition of $\sigma_{i'}$ gives birth to a homology class that dies when we add $\sigma_{j'}$. If $\sigma_{i'}$ is in the upper star of u_i, and $\sigma_{j'}$ is in the upper star of u_j, then we get (ν_i, ν_j) as the corresponding birth-death pair. It is quite possible that $i = j$, namely if both simplices belong to the same upper star, in which case we talk of a *still-birth*. We draw this birth-death pair as the point (ν_i, ν_j) in the *birth-death* plane. Alternatively, we can also draw them as $(\nu_i + \nu_j, \nu_j - \nu_i)$ in the plane. This amounts to a scaling by a factor of $\sqrt{2}$ and a rotation of coordinates by 45 degrees clock-wise. This is our preferred representation of the persistence diagrams throughout this paper. An illustration of the transformation is depicted in Figure B.2. Drawing all points representing p-dimensional homology classes gives the *p-th persistence diagram* of ϱ, which

we denote as $\mathrm{Dgm}_p(\varrho)$. Recall that the second coordinate is the persistence, and because a still-birth has zero persistence, it is drawn right on the horizontal axis. The persistence is a measure of significance of the feature represented by a birth-death point, and still-births are artifacts of the representation of ϱ and have indeed no significance. The first coordinate is the sum of birth- and death-values, and we refer to half that coordinate as the *mean age*. It gives information about the range of density values the corresponding feature is visible.[3]

Persistence diagrams contain more information than the Betti numbers. Indeed, we can read the p-th Betti number of the superlevel set for ν as the number of points of $\mathrm{Dgm}_p(\varrho)$. The contribution to the Betti numbers for a level set ν comes from all the dots in the persistence diagram corresponding to cycles that are born before ν and die after ν – in other words, the shaded region in panel (a) anchored at (ν, ν) in Figure B.2. The shaded region transforms appropriately in panel (b) to a V-shaped region anchored at $(\nu, 0)$ on the horizontal axis. The arms of the V have slope -2 and 2 respectively. Another useful property is the stability of the diagram under small perturbations of the input. Specifically, the diagram of a density function, ϱ', that differs from ϱ by at most ε at every point of the space, has bottleneck distance at most ε from $\mathrm{Dgm}_p(\varrho)$; see (Cohen-Steiner et al. 2007). This implies that every points of $\mathrm{Dgm}_p(\varrho')$ is at distance at most ε from a point in $\mathrm{Dgm}_p(\varrho)$ or from the horizontal axis.

[3] Almost every homology class that is ever born will also die at finite time, but there are eight exceptions, namely the classes that describe the 3-torus itself. They are not relevant for the study in this paper, and we do not draw them in the diagrams.

C

Stochastic Random Fields

The cosmic density perturbation field is a realization of a stochastic random field. A *random field*, f, on a spatial volume assigns a value, $f(\mathbf{x})$, to each location, $\mathbf{x}$, of that volume. The fields of interest are smooth and continuous [1]. The stochastic properties of a random field are defined by its *N-point joint probabilities*, where N can be any arbitrary positive integer. To denote them, we write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$ for a vector of N points and $\mathbf{f} = (f_1, f_2, \ldots, f_N)$ for a vector of N field values. The joint probability is

$$\text{Prob}[f(\mathbf{x}_1) = f_1, \ldots, f(\mathbf{x}_N) = f_N] \quad = \quad \mathcal{P}_{\mathbf{X}}(\mathbf{f})\,\mathrm{d}\mathbf{f}\,, \tag{C.1}$$

which is the probability that the field f at the locations $\mathbf{x}_i$ has values in the range f_i to $f_i + \mathrm{d}f_i$, for each $1 \leq i \leq N$.

In cosmological circumstances, we use the *statistical cosmological principle*, which states that statistical properties of e.g. the cosmic density distribution in the Universe are uniform throughout the Universe. It means that the distribution functions and moments of fields are the same in each direction and at each location. The latter implies that ensemble averages depend only on one parameter, namely the distance between the points.

Important for the cosmological reality is the validity of the *ergodic principle*. The Universe is unique, and its density distribution is the only realization we have of the underlying probability distribution. The ergodic principle allows us to measure the value of ensemble averages on the basis of spatial averages. These will be equal to the expectations over an ensemble of Universes, something which is of key significance for the ability to test theoretical predictions for stochastic processes like the cosmic mass distribution with observational reality.

[1] In this section, the fields $f(\mathbf{x})$ may either be the raw unfiltered field or, without loss of generality, a filtered field $f_s(\mathbf{x})$. A filtered field is a convolution with a filter kernel $W(\mathbf{x}, \mathbf{y})$, $f_s(\mathbf{x}) = \int \mathrm{d}\mathbf{y} f(\mathbf{y}) W(\mathbf{x}, \mathbf{y})$.

C.1 Gaussian Random Fields

The primordial density field is, to high accuracy, a Gaussian random field. A Gaussian random field is an example of a stochastic process. Given a parameter space T, a stochastic process f over T is a collection of random variables

$$\{f(t) : t \in T\}\,. \tag{C.2}$$

f is a (N, d) random field if T is a set of dimension N and $f(t)$ are vector-valued with dimension d. In the scope of this paper, we are interested in scalar random fields defined in $\mathbb{R}^3$. In this case, $N = 3$ and $d = 0$. Finally, f is a *gaussian* random field if $f(t)$ are gaussian distributed.

For a *Gaussian random field*, the joint probabilities for $N = 1$ and $N = 2$ determine all others (see Adler 1981; Adler & Taylor 2010; Bardeen et al. 1986). Specifically, the probability density functions take the simple form

$$\mathcal{P}_{\mathbf{X}}(\mathbf{f}) \;=\; C \cdot \exp\left[-\mathbf{f}\mathsf{M}^{-1}\mathbf{f}^T/2\right]\,, \tag{C.3}$$

where $C = 1/[(2\pi)^N(\det \mathsf{M})]^{1/2}$ normalizes the expression, making sure that the integral of $\mathcal{P}_{\mathbf{X}}(\mathbf{f})$, over all $\mathbf{f} \in \mathbb{R}^N$, is equal to 1. Here, we assume that each 1-point distribution is Gaussian with zero mean. The matrix M^{-1} is the inverse of the $N \times N$ covariance matrix with entries

$$M_{ij} \;=\; \langle f(\mathbf{x}_i) f(\mathbf{x}_j)\rangle\,, \tag{C.4}$$

in which the angle bracket denotes the ensemble average of the product, over the 2-point probability density function. In effect, M is the generalization of the variance of a 1-point normal distribution, and we indeed have $\mathsf{M} = [\sigma_0^2]$ for the case $N = 1$.

Equation (C.3) shows that a Gaussian random is fully specified by the autocorrelation function, $\xi(r)$, which expresses the correlation between the density values at two points separated by a distance $r = |\mathbf{r}|$,

$$\xi(r) \;=\; \xi(|\mathbf{r}|) \;\equiv\; \langle f(\mathbf{x}) f(\mathbf{x}+\mathbf{r})\rangle\,. \tag{C.5}$$

In other words, the entries in the matrix are the values of the *autocorrelation function* for the distance between the points: $M_{ij} = \xi(r_{ij})$, with $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$. For a multi-scale Gaussian density field, its structure can be characterized in a more transparent fashion in terms of the Fourier transform of the correlation function, the power spectrum $P(k)$. It specifies the mean square of the fluctuations of the Fourier components $\hat{f}(\mathbf{k})$ of the field $f(\mathbf{x})$,

$$f(\mathbf{x}) = \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}\, \hat{f}(\mathbf{k})\, e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{C.6}$$

via the relation

$$\langle \hat{f}(\mathbf{k})\hat{f}(\mathbf{k}')\rangle \;=\; (2\pi)^{3/2}\, P(k)\, \delta_D(\mathbf{k}-\mathbf{k}')\,, \tag{C.7}$$

where $\delta_D(\mathbf{k})$ is the Dirac delta function and the

C.1.1 Number distribution of peaks in Gaussian random fields

Bardeen et al. (1986) derive the marginal number distribution of peaks for Gaussian random fields, as a function of the dimensionless density threshold ν

$$\mathbb{N}_{pk}(\nu)\, d\nu = \frac{1}{(2\pi)^2 R_\star^3} e^{-\nu^2} G(\gamma, \gamma\nu). \tag{C.8}$$

Here, the function $G(\gamma, \gamma\nu)$ is a fitting function given by

$$G(\gamma, w) = \frac{w^3 - 3\gamma^2 w + \left[B(\gamma)w^2 + C_1(\gamma)\right] \exp\left[-A(\gamma)w^2\right]}{1 + C_2(\gamma) \exp\left[-C_3(\gamma)w\right]}. \tag{C.9}$$

The various coefficients are given by

$$\begin{aligned} A &= \frac{5/2}{9 - 5\gamma^2} \\ B &= \frac{432}{(10\pi^2)(9 - 5\gamma^2)^{5/2}} \\ C1 &= 1.84 + 1.13(1 - \gamma^2)^{5.72} \\ C2 &= 8.91 + 1.27 \exp(6.51\gamma^2) \\ C3 &= 2.58 \exp(1.05\gamma^2). \end{aligned} \tag{C.10}$$

For the power law power spectrum, the parameters γ and $R_\star$ are related to the various moments of the power spectrum, value of the spectral index n and the co-moving filtering radius R_f as

$$\begin{aligned} \gamma &= \frac{\langle k^2 \rangle}{\langle k^4 \rangle^{1/2}} \\ &= \frac{\sigma_1^2}{\sigma_2 \sigma_0} \\ &= \frac{n+3}{n+5} \\ R_\star &= \sqrt{3}\frac{\sigma_1}{\sigma_2} \\ &= \left(\frac{6}{n+5}\right)^{1/2} R_f. \end{aligned} \tag{C.11}$$

C.2 Gaussian random fields: Minkowski functionals, Euler characteristic, and genus

The goal of this sub-section is to present a brief account of the known topological characteristics of Gaussian random fields in terms of the existing decriptors in use, namely Minkowski Functionals, Euler Characteristic and genus.

C.2.1 Euler characteristic and genus

The Gauss-Bonnet Theorem relates the total Gaussian curvature κ of a connected 2-dimensional manifold surface to either their genus g or the Euler characteristic χ, suggesting that the knowledge of either κ, g or χ is sufficient to compute the others:

$$C = 4\pi(1-g) = 2\pi\chi. \tag{C.12}$$

Gott et al. (1986) derive a closed form analytical expression for the expected value of κ to arrive at expected analytic equation of g in case of Gaussian random fields

$$g(\nu) \;=\; -\frac{1}{8\pi^2}\left(\frac{\langle k^2\rangle}{3}\right)^{3/2}(1-\nu^2)e^{-\nu^2/2}, \tag{C.13}$$

where $\nu = \delta/\sigma$. Here, δ is the over- or under-density at a spatial location and σ is the rms of the density fluctuation field.
Often in cosmological literature (Doroshkevich 1970; Bardeen et al. 1986), the distinction between Euler characteristic and genus is blurred. Indeed, the expression of genus used in cosmology (Gott et al. 1986; Hamilton et al. 1986) is the same as the standard expression of Euler characteristic (Adler 1981).

C.2.2 Minkowski Functionals

There are $(d+1)$ Minkowski functionals defined for a d-dimensional manifold(Mecke et al. 1994; Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). Predominantly geometric in nature, the first four Minkowski functionals ($d \leq 3$) are respectively the volume, surface area, integrated mean curvature and the Gaussian curvature. For Gaussian random fields, the first four Minkowski functionals have known analytical epressions as a function of the superlevel sets of rms density threshold ν (Tomita 1993; Schmalzing & Buchert 1997)

$$\begin{aligned}
\mathrm{Q}_0(\nu) &= \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{1}{\sqrt{2}}\nu\right), \\
\mathrm{Q}_1(\nu) &= \frac{2}{3}\frac{\lambda}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\nu^2\right), \\
\mathrm{Q}_2(\nu) &= \frac{2}{3}\frac{\lambda^2}{\sqrt{2\pi}}\nu\exp\left(-\frac{1}{2}\nu^2\right), \\
\mathrm{Q}_3(\nu) &= \frac{\lambda^3}{\sqrt{2\pi}}(\nu^2-1)\exp\left(-\frac{1}{2}\nu^2\right).
\end{aligned} \tag{C.14}$$

where $\lambda = \sqrt{|\xi''(0)|/[2\pi\xi(0)]}$, and the function $\Phi(x) = \int_0^x \mathrm{d}te^{(-t^2)}$ is the standard error function.

The functional $Qspace_3(\nu)$ in (C.14) and the equation for genus in (C.13) differ only by a factor. This is not surprising, as the fourth Minkowski functional can also be formulated as an integral of the Gaussian curvature, which is related to Euler characteristic and genus via the Gauss-Bonneť Theorem. In fact, Crofton's intersection formula (Crofton 1868) goes further and establishes that all the Minkowski functionals can be formulated as appropriate integrals of Euler characteristic (Pranav et al. 2013). Evidently, Euler characteristic is the more fundamental quantity, and is the bridge between geometry and topology (Adler & Taylor 2009).

D

Voronoi Clustering Models

Voronoi clustering models are heuristic models for cellular spatial patterns which use the geometric (and convex) structure of the Voronoi tessellation (Voronoi 1908; Okabe et al. 2000) to emulate the cosmic matter distribution van de Weygaert & Icke (1989); van de Weygaert (1991, 2002). They offer flexible templates for cellular patterns and are easy to tune towards a specific spatial cellular morphology. This makes them very suitable for studying clustering properties of nontrivial geometric spatial patterns. Unless otherwise specified, the seeds of the tessellation usually involve a set of Poisson distributed points.

The Voronoi models use Voronoi tessellations for defining the structural frame around which matter assembles as cosmic structure emerges and grows. Particles are distributed within this skeleton by assigning them to one of the four distinct structural components of a Voronoi tessellation. The interior of Voronoi cells is identified with *void regions*, the Voronoi cell faces with *walls*, the edges with *filaments* and the vertices with *cluster nodes*. What is usually described as a flattened "supercluster" consists of an assembly of various connecting walls in the Voronoi foam, while elongated "superclusters" of "filaments" usually include a few coupled edges. Vertices are the most outstanding structural elements, corresponding to the very dense compact nodes within the cosmic web where one finds the rich clusters of galaxies.

Among a variety of possible Voronoi clustering realizations, two distinct yet complementary classes of models are the most frequently used ones, the structurally rigid *Voronoi Element Models* and the evolving *Voronoi evolution models*. Both the Voronoi Element Models and the Voronoi Evolution Models are obtained by projecting an initially random distribution of N sample points/galaxies onto the walls, edges or vertices of the Voronoi tessellation defined by M nuclei.

D.0.3 Voronoi element models

"Voronoi element models" are fully heuristic models. They are user-specified spatial galaxy distributions within the *cells* (field), *walls*, *edges* and *vertices* of a Voronoi tessellation. *Pure Voronoi element Models* place their model galaxies exclusively in either walls, edges or vertices.

The practical implementation of the Voronoi model consists of an initial random distribution of N particles in a box of volume V with periodic boundary conditions (ie. a 3-torus). The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\mathbf{x}_{n0}$ $(n = 1, \ldots, N)$ defined by a homogeneous Poisson process. A random distribution of M nuclei defines the geometric structure of the Voronoi tessellation.

The initially randomly distributed model particles are projected onto the relevant Voronoi wall, Voronoi edge or Voronoi vertex or retained within the interior of the Voronoi cell in which they are located, according to a process which is the asymptotic limit of the prescription outlined in equation D.2 (see van de Weygaert et al. 2011). The model walls, filaments and nodes are not infinitely thin, but have a Gaussian density profile with a user-specified width. Following their projection onto wall, filament or vertex, the particles are randomly displaced according to this profile.

The versatility of the Voronoi element model also allows combinations in which field (cell), wall, filament and vertex distributions are superimposed. These complete composite particle distributions, *Mixed Voronoi element Models*, include particles located in four distinct structural components. The characteristics of the patterns and spatial distribution in the composite Voronoi Element models can be varied and tuned according to the fractions of galaxies in in Voronoi walls, in Voronoi edges, in Voronoi vertices and in the field. These fractions are free parameters to be specified by the user.

D.0.4 Voronoi evolution models

The second class of Voronoi models is that of the *Voronoi Evolution models*. They attempt to provide weblike galaxy distributions that reflect the outcome of realistic cosmic structure formation scenarios. They are based upon the notion that voids play a key organizational role in the development of structure and makes the Universe resemble a soapsud of expanding bubbles Icke (1984). While the galaxies move away from the void centres, and stream out of the voids towards the sheets, filaments and clusters in the Voronoi network the fraction of galaxies in the voids (cell interior), the sheets (cell walls), filaments (wall edges) and clusters (vertices) is continuously changing and evolving. The details of the model realization depends on the time evolution specified by the particular Voronoi Evolution Model.

Within the class of Voronoi Evolution Models the most representative and most frequently used are the *Voronoi kinematic models*. They form the idealized and asymptotic description of the outcome of hierarchical gravitational structure formation process, with single-sized voids forming around depressions in the primordial density field. This is translated into a scheme for the displacement of initially randomly distributed galaxies within the Voronoi skeleton. Within a void, the mean distance between galaxies increases uniformly in the course of time. When a galaxy tries to

enter an adjacent cell, the velocity component perpendicular to the cell wall disappears. Thereafter, the galaxy continues to move within the wall, until it tries to enter the next cell; it then loses its velocity component towards that cell, so that the galaxy continues along a filament. Finally, it comes to rest in a node, as soon as it tries to enter a fourth neighbouring void.

As for the Voronoi element models, the Voronoi kinematic models are based on the displacement of a sample of N particles, or model galaxies, with an initial random distribution. In practice, these are distributed in a box of volume V with periodic boundary conditions (ie. a 3-torus). The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\mathbf{x}_{n0}$ $(n = 1, \ldots, N)$ defined by a homogeneous Poisson process. A set of M nuclei or *expansion centres* within the volume V corresponds to the cell centres, or "expansion centres" driving the evolving matter distribution. The nuclei have locations $\mathbf{y}_m$ $(m = 1, \ldots, M)$. The first step of the formalism is to determine for each galaxy n the Voronoi cell $\mathcal{V}_\alpha$ in which it is initially located.

The path $\mathbf{x}_n$ along which a galaxy moves within the Voronoi skeleton depends on how far it has moved away from its initial location $\mathbf{x}_{n0}$. For a specific galaxy n this path may consist of the following sequence,

- cell displacement $\mathbf{s}_\alpha$ radially directed away from the expansion centre j_α
- wall displacement $\mathbf{s}_{\alpha\beta}$ within the Voronoi wall $\Sigma_{\alpha\beta}$, $\Sigma_{\alpha\beta}$ defined by the nucleus j_α and its natural neighbour j_β
- edge displacement $\mathbf{s}_{\alpha\beta\gamma}$ along the Voronoi edge $\Lambda_{\alpha\beta\gamma}$, $\Lambda_{\alpha\beta\gamma}$ defined by j_α and its natural neighbours j_β and j_γ .

This path is encapsulated in the equation,

$$\begin{aligned} \mathbf{x}_n &= \mathbf{y}_\alpha + \mathbf{s}_{n\alpha} + \mathbf{s}_{n\alpha\beta} + \mathbf{s}_{n\alpha\beta\gamma} \\ &= \mathbf{y}_\alpha + s_{n\alpha}\hat{\mathbf{e}}_{n\alpha} + s_{n\alpha\beta}\hat{\mathbf{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}\hat{\mathbf{e}}_{n\alpha\beta\gamma}\,, \end{aligned} \tag{D.1}$$

where the unity vectors $\hat{\mathbf{e}}_{n\alpha}$, $\hat{\mathbf{e}}_{n\alpha\beta}$ and $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$ specify the direction of galaxy's path within a Voronoi cell, a Voronoi wall or Voronoi edge.

While the galaxy's location in the Voronoi element models (see section D.0.3) is restricted to one of the components of the Voronoi skeleton, the Voronoi kinematic model is characterized by an evolving global "void" expansion factor $R(t)$ which dictates whether a galaxy has entered a wall, proceeded towards a filament or has arrived within a vertex/cluster node. At any one cosmic epoch each galaxy n is displaced according to $R(t)$. At first, while still within the cell's interior, the galaxy proceeds radially from its expansion centre j_α, moving along the radial cell path emanating from its nucleus,

$$\begin{aligned} s_{n\alpha}(t) &= \mathbf{x}_n(t) - \mathbf{y}_\alpha \\ &= R(t)\,|\mathbf{x}_{n0} - \mathbf{y}_\alpha|\,\hat{\mathbf{e}}_{n\alpha}\,. \end{aligned} \tag{D.2}$$

During this stage the wall and edge path factors $s_{n\alpha\beta} = s_{n\alpha\beta\gamma} = 0$, while the mean distance between the particles increases uniformly with expansion factor $R(t)$. Once the galaxy tries to enter an adjacent cell j_β and reaches a Voronoi wall, i.e. when

$$R(t)\,|\mathbf{x}_{n0} - \mathbf{y}_\alpha| \;>\; v_n \tag{D.3}$$

the gravity of the wall, aided and abetted by dissipational processes, will slow down its motion. Subsequently, the velocity component perpendicular to the cell wall disappears and the galaxy continues to move within the wall along the direction of $\hat{\mathbf{e}}_{n\alpha\beta}$. This continues until it reaches the corresponding Voronoi edge, after which it moves along the direction $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$. Finally, it will reach the Voronoi vertex towards its move, which represents the cluster node where it will culminate its cosmic journey. As in the case of the Voronoi element models, the walls, filaments and nodes in the Voronoi kinematic models have a Gaussian density profile, with a user-specified width. To effect this profile, following their arrival in wall, filament or vertex node, the particles are randomly displaced according to the specified profile. For further details of the complete formalism for generating these spatial distributions can be found in van de Weygaert et al. (2011).

The resulting evolutionary progression within the Voronoi kinematic scheme is that of an almost featureless random distribution, via a wall-like and filamentary morphology towards a distribution in which matter ultimately aggregates into conspicuous compact cluster-like clumps. Figure 2.16 provides an impression of the evolutionary progression.

E

Soneira-Peebles model.

The Soneira-Peebles model is an analytic self-similar spatial point distribution which was defined for the purpose of modelling the galaxy distribution, such that its statistical properties would be tuned to reality (Soneira & Peebles 1978). An important property of the Soneira-Peebles model is that it is one of the few nonlinear models of the galaxy distribution whose statistical properties can be fully and analytically evaluated. This concerns its power-law two-point correlation function, correlation dimension and its Hausdorff dimension. Here we shortly specify the main characteristics of the Soneira-Peebles model, for an extensive description of the fractal-like properties of the Soneira-Peebles model we refer to (Martinez 1990) and (van de Weygaert & Schaap 2009b).

The Soneira-Peebles model is specified by three parameters (for an illustrated descriptions see van de Weygaert & Schaap (2009b)). The starting point of the model is a level-0 sphere of radius R. At each level-m a number of ψ subspheres are placed randomly within their parent level-m sphere: the level-$(m+1)$ spheres have a radius R/ζ where $\zeta > 1$, the size ratio between parent sphere and subsphere. This process is repeated for L successive levels, yielding ψ^L level-L spheres of radius R/ζ^L. At the center of each of these spheres a point is placed, producing a point sample of ζ^L points. While this produces a pure *singular* Soneira-Peebles model, usually a set of these is superimposed to produce a somewhat more realistically looking model of the galaxy distribution, an *extended* Soneira-Peebles model.

E.0.5 Self-similarity

The Soneira-Peebles model involves a hierarchy of structures of varying densities and characteristic scales, with the higher level spheres corresponding to high density structures of small scale and the lower level spheres corresponding to low density

structures of larger scale. As each sphere is constructed in the same way, the resulting point distribution is self-similar and forms a bound fractal. The fractal geometry of a point set is often characterized by the fractal dimension D, which is defined as

$$D = \lim_{r \to 0} \frac{\log N(r)}{\log(1/r)} \tag{E.1}$$

Here $N(r)$ is the number of non-empty cells in a partition of constant cell size r. If the Soneira-Peebles model would contain an infinite amount of levels, the resulting point distribution would have fractal dimension

$$D = (\log \psi)/(\log \zeta)\,. \tag{E.2}$$

One important manifestation of the self-similarity of the defined Soneira-Peebles distribution is reflected in the power-law two-point correlation function. For M dimensions it is given by

$$\begin{aligned} \xi(r) &\sim r^{-\gamma}\,, \\ \gamma &= M - \left(\frac{\log \psi}{\log \zeta}\right) \quad \text{for} \quad \frac{R}{\zeta^{L-1}} < r < R\,. \end{aligned} \tag{E.3}$$

The parameters ψ and ζ may be adjusted such that they yield the desired value for the correlation slope γ.

F

Density Estimators

In the cosmological reality, a density field is discretely sampled by galaxies, or by particles in computer simulations. This discrete distribution is subsequently translated into a continuous density field, based on the key assumption that the discrete point distribution represents a fair discrete sampling of the underlying density field. There is a large array of techniques available for obtaining density estimates from a discrete particle distribution. Key references on the corresponding problems and solutions include those by Ripley (1981). An extensive and systematic survey of available mathematical methods in an astronomical context can be found in a set of publications by Lombardi & Schneider (2001); Lombardi (2002); Lombardi & Schneider (2003).

In this appendix we focus on two formalisms. This study is mainly based on the use of the Delaunay Tessellation Field Estimator (DTFE). In addition, we have compared the obtained topological results with that of a spatially adaptive kernel estimator used in Smooth Particle Hydrodynamics.

F.0.6 DTFE, the Delaunay Tessellation Field Estimator

DTFE - the Delaunay Tessellation Field Estimator - exploits three properties of Voronoi and Delaunay tessellations. The first is the sensitivity of the tessellation cell size to the local point density. The DTFE method uses this fact to define a local estimate of the density, on the basis of the inverse of the volume of the tessellation cells. Equally important is their sensitivity to the local shape of the point distribution, which allows them to trace anisotropic features. Finally, it uses the adaptive and minimum triangulation properties of Delaunay tessellations to use them as adaptive spatial interpolation intervals for irregular point distributions.

In the first step of the DTFE procedure, a local density estimate ϱ_i is computed at each point i of the point sample. The value is the (normalized) inverse of the volume $V(\mathcal{W}_i)$ of the corresponding *contiguous Voronoi cell* or *star*, $\mathcal{W}_i$. The star consists of all Delaunay simplices that contain $\vec{x}_i$ as a vertex. In D dimensions the DTFE density estimate at each sample point i is given by:

$$\varrho_i \;=\; \frac{(1+D)}{V(\mathcal{W}_i)}\,. \tag{F.1}$$

DTFE subsequently interpolates the values of these density field estimates $f_i = \varrho_i$ over the volume of the sample. This produces a piecewise linear interpolation of the field, in which the Delaunay tessellation is used as interpolation grid. The field gradient ∇f is defined to be constant over the volume of each Delaunay tetrahedron, with its value directly and uniquely determined from the $(1+D)$ field values f_j at the sample points constituting the vertices of a Delaunay simplex.

Given the location $\vec{r}_0$, $\vec{r}_1$, $\vec{r}_2$ and $\vec{r}_3$, of the four points forming the Delaunay tetrahedra's vertices, and the field value estimates at these locations, f_0, f_1, f_2 and f_3, the gradient $\widehat{\nabla} f$ follows from the inversion

$$\widehat{\nabla} f \;=\; \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2ex] \dfrac{\partial f}{\partial y} \\[2ex] \dfrac{\partial f}{\partial z} \end{pmatrix} \;=\; \mathbf{A}^{-1} \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \Delta f_3 \end{pmatrix} \;; \qquad \mathbf{A} \;=\; \begin{pmatrix} \Delta x_1 & \Delta y_1 & \Delta z_1 \\ \Delta x_2 & \Delta y_2 & \Delta z_2 \\ \Delta x_3 & \Delta y_3 & \Delta z_3 \end{pmatrix} \tag{F.2}$$

where the coordinate distance between the tetrahedral vertices is represented by $\Delta x_n = x_n - x_0$, $\Delta y_n = y_n - y_0$ and $\Delta z_n = z_n - z_0$, and the corresponding differences between the vertex field values is given by $\Delta f_n \equiv f_n - f_0\,(n = 1, 2, 3)$.

Once the value of ∇_f for each Delaunay tetrahedron has been determined, it is straightforward to determine the DTFE field value $\widehat{f}(\vec{x})$ for any location $\vec{x}$ by means of straightforward linear interpolation within the Delaunay tetrahedron m in which $\vec{x}$ is located (eqn. F.2),

$$\widehat{f}(\vec{x}) \;=\; \widehat{f}(\vec{x}_i) \;+\; \widehat{\nabla f}\big|_m \cdot (\vec{x} - \vec{x}_i)\,. \tag{F.3}$$

In many practical situations we need a DTFE reconstructed field sampled on a grid. The created DTFE image may be then be used for further analysis. To this end, the DTFE density field value is sampled at each gridpoint of the grid. Formally, this should involve the DTFE density field value averaged over the corresponding gridcell. That value could be determined geometrically, although a more practical implementation involves the averaging over the interpolated field values (eqn. F.3) over a number or randomly placed points in a gridcell (e.g. Cautun et al. 2013). In a number of practical circumstances, a reasonable shortcut is to limit the field value calculation to that at the grid location. This offers a reasonable approximation for

gridcells which are smaller or comparable to that of intersecting Delaunay cells, on the condition that the field gradient within the cell(s) is not too large.

F.0.7 SPH kernel density estimates

In the previous appendix, we have described in some detail the technical aspects of DTFE. DTFE produces volume-weighted density estimates that adapt themselves to the local number density and geometry/shape of the particle distribution. The majority of conventionally applied techniques produce mass-weighted estimates. They usually involve a suitably weighted sum over discretely sampled field values, involving kernel weight functions $W(\mathbf{x}, \mathbf{y})$. One may directly appreciate that such convolutions produce mass-weighted estimates by converting the discrete sum into an integral over Dirac delta functions (see van de Weygaert & Schaap 2009b),

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \frac{\sum_{i=1}^{N} \tilde{f}_i \, W(\mathbf{x}-\mathbf{x}_i)}{\sum_{i=1}^{N} W(\mathbf{x}-\mathbf{x}_i)} \\ &= \frac{\int d\mathbf{y}\, f(\mathbf{y})\, W(\mathbf{x}-\mathbf{y}) \, \sum_{i=1}^{N} \delta_D(\mathbf{y}-\mathbf{x}_i)}{\int d\mathbf{y}\, W(\mathbf{x}-\mathbf{y}) \, \sum_{i=1}^{N} \delta_D(\mathbf{y}-\mathbf{x}_i)} \\ &= \frac{\int d\mathbf{y}\, f(\mathbf{y})\, \rho(\mathbf{y})\, W(\mathbf{x}-\mathbf{y})}{\int d\mathbf{y}\, \rho(\mathbf{y})\, W(\mathbf{x}-\mathbf{y})} . \end{aligned} \tag{F.4}$$

Both rigid grid-based convolution schemes as well as a convolution formalism involving a spatially adaptive technique produce a mass-weighted average.

One particularly interesting example of the latter are the adaptive density functions used in Smooth Particle Hydrodynamics (SPH) codes (Hernquist & Katz 1989; Monaghan 2005; Springel 2008, see e.g.). The basic feature of the SPH procedure for density estimation consist of the convolution of the discrete particle distribution with a user-specified kernel function W that adapts itself to the local point density. For a sample of N particles, with masses m_j and locations $\mathbf{r}_j$, the density estimate ρ_{SPH} at the location $\mathbf{r}_i$ of particle i is given by

$$\rho(\mathbf{r}_i) \, = \, \sum_{j=1}^{N} m_j \, W(\mathbf{r}_i - \mathbf{r}_j, h_i) \, , \tag{F.5}$$

in which the kernel resolution is determined through the smoothing scale h_i. In general, the scale h_i is set by the local particle density. In many applications, the smoothing length h_i is chosen such that the sum involves a specific number of M nearest neighbours, usually in the order of $M \approx 40$.

The functional dependence of the kernel W is nearly always spherically summetric, so that it is the function of $|\mathbf{r}_i - \mathbf{r}_j|$ only. One possibility would be the use of Gaussian kernels W whose width h would be set by the local number density. More elaborate schemes, such as described by Monaghan (2005), use spline kernels. A

more advanced SPH scheme, introduced by Hernquist & Katz (1989), uses a symmetrized form of Eq. F.5. In general, the variants of SPH density estimates produce comparable results.

G

Skew-normal distribution

The skew-normal distribution with the *skewness parameter* α is given by(O'Hagan & Leonard (1976); Azzalini (1985))

$$f(x) = 2\phi(\nu)\Phi(\alpha\nu). \tag{G.1}$$

Here,

$$\phi(\nu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\nu^2}{2}}. \tag{G.2}$$

is the standard normal distribution. The function

$$\begin{aligned} \Phi(\nu) &= \int_{-\infty}^{x} \phi(t)\, dt \\ &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\nu}{\sqrt{2}}\right)\right], \end{aligned} \tag{G.3}$$

is the cumulative distribution function, in which *erf(x)* is the error function. Note that one recovers the familiar normal distribution when the skewness parameter $\alpha = 0$. he absolute value of skewness increases as the absolute value of α increases. Note that the skewness parameter α is different than the skewness, i.e. the third moment of the distribution

$$\gamma_1 = \mu^3/\sigma^3. \tag{G.4}$$

By definition, a curve has positive skewness if it has a more prominent tail for increasing values of ν, and a negative skewness when its balance is shifted towards decreasing values of ν.

To account for the location of the peak and the width of the curve, one usually makes the transformation

$$\mathrm{x} \to \mathrm{x} - \xi/\omega, \tag{G.5}$$

where ξ and ω are the location and scale parameters respectively.
The probability distribution function with location ξ, scale ω and skewness parameter α becomes

$$f(x) = \frac{2}{\omega}\phi\left(\frac{\nu-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{\nu-\xi}{\omega}\right)\right). \tag{G.6}$$

Introducing the amplitude parameter A_0, this takes the form

$$f(\nu) = \frac{A_0}{\omega\pi}\mathrm{e}^{-\frac{(\nu-\xi)^2}{2\omega^2}}\int_{-\infty}^{\alpha\left(\frac{\nu-\xi}{\omega}\right)}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t. \tag{G.7}$$

Bibliography

Aarseth, S. J. 1972, in Astrophysics and Space Science Library, Vol. 31, IAU Colloq. 10: Gravitational N-Body Problem, ed. M. Lecar, 373

Aarseth, S. J. 1973, in Astrophysics and Space Science Library, Vol. 39, Recent Advances in Dynamical Astronomy, ed. B. D. Tapley & V. Szebehely, 197

Aarseth, S. J. 2011, NBODY Codes: Numerical Simulations of Many-body (N-body) Gravitational Interactions, astrophysics Source Code Library

Aarseth, S. J. & Dekel, A. 1983, in Clustering in the Universe, ed. D. Gerbal & A. Mazure, 13

Aarseth, S. J., Turner, E. L., & Gott, III, J. R. 1979, The Astrophysical Journal, 228, 664

Abazajian, K., Adelman-McCarthy, J. K., & Agüeros, M. A. e. a. 2003, Astrophysical Jornal, 126, 2081

Abel, T., Anninos, P., Norman, M. L., & Zhang, Y. 1998, The Astrophysical Journal, 508, 518

Abel, T., Hahn, O., & Kaehler, R. 2012, MNRAS, 427, 61

Adler, R. 1981, The Geometry of Random Fields, Classics in applied mathematics (Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104))

Adler, R., Bobrowski, O., & Weinberger, S. 2014, Discrete & Computational Geometry, 52, 680

Adler, R. & Taylor, J. 2009, Random Fields and Geometry, Springer Monographs in Mathematics (Springer)

Adler, R. & Taylor, J. 2010, Random Fields and Geometry, Springer Monographs in Mathematics (Springer)

Adler, R. J., Bobrowski, O., Borman, M. S., Subag, E., & Weinberger, S. 2010, ArXiv e-prints

Alpaslan, M., Galaxy, & Mass Assembly (GAMA) survey team. 2015, in American Astronomical Society Meeting Abstracts, Vol. 225, American Astronomical Society Meeting Abstracts, 405.05

Alpaslan, M., Robotham, A. S. G., et al. 2014, MNRAS, 440, L106
Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, Mon. Not. Royal Astro. Soc., 425, 2722
Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007a, Astrophys. J. Lett., 655, L5
Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007b, Astron. Astrophys., 474, 315
Aragón-Calvo, M. A., Platen, E., et al. 2010, Astrophys. J., 723, 364
Aragon-Calvo, M. A. & Szalay, A. S. 2013, MNRAS, 428, 3409
Aragon-Calvo, M. A. & Szalay, A. S. 2013, Mon. Not. Royal Astro. Soc., 428, 3409
Aragon-Calvo, M. A., van de Weygaert, R., Araya-Melo, P. A., Platen, E., & Szalay, A. S. 2010, Mon. Not. Royal Astro. Soc., 404, L89
Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010, MNRAS, 408, 2163
Aragon-Calvo, M. A. & Yang, L. F. 2014, Mon. Not. Royal Astro. Soc., 440, L46
Azzalini, A. 1985, Scandinavian Journal of Statistics, 12, pp. 171
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, Astrophys.J., 304, 15
Barrow, J. D., Bhavsar, S. P., & Sonoda, D. H. 1985, MNRAS, 216
Bartolo, N., Matarrese, S., & Riotto, A. 2005, Jour. Cos. and Part. Phys., 10, 10
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013, The Astrophysical Journal, 763, 18
Bendich, P., Edelsbrunner, H., & Kerber, M. 2010, IEEE Transactions on Visualization and Computer Graphics, 16, 1251
Bennett, C. L., Halpern, M., Hinshaw, G., & et. al. 2003, Astrophys. J. Suppl., 148, 1
Bentley, J. L. 1975, Commun. ACM, 18
Bernardeau, F. & van de Weygaert, R. 1996, MNRAS, 279, 693
Bertschinger, E. 1987, ApJL, 323, L103
Betti, E. 1871, Ann. Mat. Pura Appl., 2, 140
Blum, H. 1967, Models for the Perception of Speech and Visual Form
Bobrowski, O. & Borman, M. S. 2012, Journal of Topology and Analysis, 4
Bobrowski, O. & Kahle, M. 2014, ArXiv e-prints
Bobrowski, O. & Mukherjee, S. 2015, Probability Theory and Related Fields, 161, 651
Bobrowski, O., Mukherjee, S., & Taylor, J. E. 2014, ArXiv e-prints
Bobrowski, O. & Strom Borman, M. 2010, ArXiv e-prints
Bok, B. J. 1934, Harvard College Observatory Bulletin, 895, 1
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440
Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603
Bond, J. R. & Myers, S. T. 1996a, The Astrophysical Journal Supplements, 103, 1
Bond, J. R. & Myers, S. T. 1996b, The Astrophysical Journal Supplements, 103, 41
Bond, J. R. & Myers, S. T. 1996c, The Astrophysical Journal Supplements, 103, 63
Bond, N. A., Strauss, M. A., & Cen, R. 2010, MNRAS, 409, 156
Bonnet, O. 1848, J. cole Polytechnique, 19, 1
Bonometto, S. A. 1995, in Large Scale Structure in the Universe, ed. J. P. Mücket, S. Gottloeber, & V. Müller, 142
Borgani, S. 1993, Mon. Not. Royal Astro. Soc., 260, 537
Bos, E. G. P., van de Weygaert, R., Dolag, K., & Pettorino, V. 2012, Mon. Not. of Royal

Astro. Soc., 426, 440
Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, Mon. Not. Royal Astro. Soc., 398, 1150
Bubenik, P. 2012, CoRR, abs/1207.6437
Canavezes, A. & Efstathiou, G. 2004, Astrophysics and Space Science, 290, 215
Canavezes, A., Springel, V., Oliver, S. J., et al. 1998, MNRAS, 297, 777
Carlsson, G. 2009, Bulletin of the American Mathematical Society, 46, 255
Carlsson, G. & Zomorodian, A. 2009, Discrete & Computational Geometry, 42, 71
Carlsson, G., Zomorodian, A., Collins, A., & Guibas, L. J. 2005, International Journal of Shape Modeling, 11, 149
Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, Mon. Not. Royal Astro. Soc., 429, 1286
Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, ArXiv e-prints
Cautun, M. C. & van de Weygaert, R. 2011, The DTFE public software: The Delaunay Tessellation Field Estimator code, astrophysics Source Code Library
Chan, K. C., Hamaus, N., & Desjacques, V. 2014, Phys. Rev. D, 90, 103521
Chazal, F., Cohen-Steiner, D., Guibas, L., Mmoli, F., & Oudot, S. 2009, Computer Graphics Forum, 28, 13931403
Chazal, F., Cohen-Steiner, D., & Mrigot, Q. 2011, Foundations of Computational Mathematics, 11, 733
Chazal, F., Fasy, B. T., Lecci, F., et al. 2013, ArXiv e-prints
Choi, Y.-Y., Park, C., Kim, J., et al. 2010, ApJ Suppl., 190, 181
Clampitt, J., Cai, Y.-C., & Li, B. 2013, MNRAS, 431, 749
Clampitt, J., Jain, B., & Khoury, J. 2012, Jour. Cos. and Part. Phys., 1, 30
Codis, S., Pichon, C., et al. 2012, MNRAS, 427, 3320
Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. 2007, Discrete Comput. Geom., 37, 103
Colberg, J. M. 2007, Mon. Not. Royal Astro. Soc., 375, 337
Cole, S. & Lacey, C. 1996, Mon. Not. Royal Astro. Soc., 281, 716
Coles, P. & Jones, B. 1991, Mon. Not. Royal Astro. Soc., 248, 1
Colless, M., Peterson, B. A., & Jackson, C. e. a. 2003, ArXiv Astrophysics e-prints
Crofton, M. W. 1868, Philosophical Transactions of the Royal Society of London, 158, 181
Davis, M. & Peebles, P. J. E. 1983, The Astrophysical Journal, 267, 465
De Boni, C., Dolag, K., Ettori, S., et al. 2011, MNRAS, 415, 2758
de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, The Astrophysical Journal, 302, L1
Dietrich, J. P., Werner, N., et al. 2012, Nature, 487, 202
Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, MNRAS, 399, 497
Dominik, K. G. & Shandarin, S. 1992, Astrophys.J., 393, 450
Doroshkevich, A. G. 1970, Astrophysics, 6, 320
Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, Astrophysical Journal Suppl., 180, 306
Durrer, R., Novosyadlyj, B., & Apunevych, S. 2003, ApJ, 583, 33
Edelsbrunner, H. 2001, Geometry and Topology for Mesh Generation, Cambridge

Monographs on Applied and Computational Mathematics (Cambridge University Press)
Edelsbrunner, H. & Harer, J. 2010, Computational Topology: An Introduction, Applied mathematics (American Mathematical Society)
Edelsbrunner, H., Harer, J., Natarajan, V., & Pascucci, V. 2003, in Proc. of Symp. on Comp. Geom., 361–370
Edelsbrunner, H., Harer, J., & Zomorodian, A. 2001, in Proc. of Symp. on Comp. Geom., 70–79
Edelsbrunner, H., Letscher, J., & Zomorodian, A. 2002, Discrete & Computational Geometry, 28, 511
Edelsbrunner, H. & Mücke, E. P. 1994, ACM Trans. Graph., 13, 43
Efstathiou, G. & Bond, J. R. 1999, MNRAS, 304, 75
Efstathiou, G. & Jones, B. J. T. 1979, Mon. Not. Royal Astro. Soc., 186, 133
Einstein, A. 1916, Annalen der Physik, 354, 769
Eisenstein, D. J. & Hu, W. 1999, ApJ, 511, 5
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, The Astrophysical Journal, 633, 560
Euler, L. 1758, Novi Commentarii academiae scientiarum Petropolitanae, 4, 140
Fairall, A. P., Turner, D., Pretorius, M. L., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance, ed. A. P. Fairall & P. A. Woudt, 229
Fasy, B. T., Lecci, F., Rinaldo, A., et al. 2013, ArXiv e-prints
Feldbrugge, J., Engelen, M. V., van de Weygaert, R., et al. 2015, to be submitted
Fixsen, D. J., Cheng, E. S., Cottingham, D. A., et al. 1994, The Astrophysical Journal, 420, 445
Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, MNRAS, 396, 1815
Forman, R. 2002, Séminaire Lotharingien de Combinatoire, 48
Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377
Gauss, C. F. 1900, K. Gesellschaft Wissenschaft, 8
Genovese, C., Perone-Pacifico, M., Verdinelli, I., & Wasserman, L. 2010, ArXiv e-prints
Gott, J. R., Choi, Y.-Y., Park, C., & Kim, J. 2009, ApJ Letters, 695, L45
Gott, III, J. R., Dickinson, M., & Melott, A. L. 1986, Astrophysical Journal, 306, 341
Gott, III, J. R., Hambrick, D. C., Vogeley, M. S., et al. 2008, The Astrophysical Journal, 675, 16
Gott, III, J. R., Miller, J., Thuan, T. X., et al. 1989, Astrophysical Journal, 340, 625
Gott, III, J. R., Park, C., Juszkiewicz, R., et al. 1990, Astrophysical Journal, 352, 1
Grainge, K., Carreira, P., Cleary, K., Davies, R. D., & et. al. 2003, MNRAS, 341, L23
Groth, E. J. & Peebles, P. J. E. 1977, The Astrophysical Journal, 217, 385
Groth, E. J. & Peebles, P. J. E. 1986, The Astrophysical Journal, 310, 507
Gunn, J. E. & Gott, III, J. R. 1972, The Astrophysical Journal, 176, 1
Günther, D., Reininghaus, J., Seidel, H. P., & Weinkauf, T. 2014, in Topological Methods in Data Analysis and Visualization III, ed. P.-T. Bremer, I. Hotz, V. Pascucci, & R. Peikert, Mathematics and Visualization (Springer International Publishing)

Günther, D., Seidel, H.-P., & Weinkauf, T. 2012, Comp. Graph. Forum, 31, 2554
Gurbatov, S. N., Saichev, A. I., & Shandarin, S. F. 1989, Mon. Not. Royal Astro. Soc., 236, 385
Guth, A. H. 1981, Physical Review D, 23, 347
Guth, A. H. & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110
Guzzo, L. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 94, Mapping, Measuring, and Modelling the Universe, ed. P. Coles, V. Martinez, & M.-J. Pons-Borderia, 157
Guzzo, L. & The Vipers Team. 2013, The Messenger, 151, 41
Gyulassy, A., Bremer, P.-T., et al. 2008, IEEE Trans. Vis. & Comp. Graphics, 14, 1619
Gyulassy, A., Kotava, N., et al. 2012, IEEE Trans. Vis. & Comp. Graphics, 18, 1549
Gyulassy, A., Natarajan, V., et al. 2007, IEEE Trans. Vis. & Comp. Graphics, 13, 1440
Gyulassy, A. G. 2008, PhD Thesis
Hadwiger, H. 1957, Vorlesungen über Inhalt, Oberfläche und Isoperimetrie, Grundlehren der mathematischen Wissenschaften (Springer)
Hahn, O., Porciani, C., et al. 2007, MNRAS, 375, 489
Hamilton, A. J. S. 1993, The Astrophysical Journal, 417, 19
Hamilton, A. J. S., Gott, III, J. R., & Weinberg, D. 1986, Astrophysical Journal, 309, 1
Harrison, E. R. 1970, Phys. Rev. D, 1, 2726
Hazra, D. K., Shafieloo, A., & Souradeep, T. 2013, JCAP, 7, 31
Heavens, A. & Peacock, J. 1988, Mon. Not. Royal Astro. Soc., 232, 339
Hellwing, W. A., Barreira, A., Frenk, C. S., Li, B., & Cole, S. 2014, Physical Review Letters, 112, 221102
Helmi, A. & White, S. D. M. 1999, Mon. Not. Royal Astro. Soc., 307, 495
Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. 1999, Nature, 402, 53
Hernquist, L. & Katz, N. 1989, Apj Suppl., 70, 419
Hidding, J., Shandarin, S. F., & van de Weygaert, R. 2014, Mon. Not. Royal Astro. Soc., 437, 3442
Hidding, J., van de Weygaert, R., Vegter, G., & Jones, B. J. T. 2015, in Thirteenth Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories, ed. K. Rosquist, 2142–2144
Hikage, C., Coles, P., Grossi, M., et al. 2008, Mon. Not. Royal Astro. Soc., 385, 1613
Hikage, C., Komatsu, E., & Matsubara, T. 2006, The Astrophysical Journal, 653, 11
Hikage, C., Suto, Y., Kayo, I., et al. 2002, Pub. Astro. Soc. of Japan, 54, 707
Hoeflich, P. & Khokhlov, A. 1996, The Astrophysical Journal, 457, 500
Houlahan, P. & Scalo, J. 1992, The Astrophysical Journal, 393, 172
Hoyle, F. 1951a, in Problems of Cosmical Aerodynamics, 195
Hoyle, F. 1951b, in Problems of Cosmical Aerodynamics, ed. J. M. Burgers & H. C. van de Hulst, 195
Hu, W. & Eisenstein, D. J. 1999, Phys. Rev. D, 59, 083509
Huchra, J., Jarrett, T., Skrutskie, M., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance, ed. A. P. Fairall & P. A. Woudt, 135
Huchra, J. P. & Geller, M. J. 1982, The Astrophysical Journal, 257, 423

Icke, V. 1973, Astron. & Astrophysics, 27, 1
Icke, V. 1984, MNRAS, 206, 1P
Icke, V. & van de Weygaert, R. 1991, Quarterly Journal of the Roy. Astro. Soc., 32, 85
Ishiyama, T., Rieder, S., et al. 2013, Astrophys. J., 767
James, J. B., Colless, M., Lewis, G. F., & Peacock, J. A. 2009, MNRAS, 394, 454
Jennings, E., Li, Y., & Hu, W. 2013, Mon. Not. Royal Astro. Soc., 434, 2167
Jiang, F. & van den Bosch, F. C. 2014, Mon. Not. Royal Astro. Soc., 440, 193
Jones, B. J. T. 1992, The large scale structure of the Universe, ed. F. Sanchez, M. Collados, & R. Reboldo, 171
Jones, B. J. T. & Efstathiou, G. 1979, Mon. Not. Royal Astro. Soc., 189, 27
Jones, B. J. T., van de Weygaert, R., & Aragón-Calvo, M. A. 2010, Mon. Not. Royal Astro. Soc., 408, 897
Kamionkowski, M., Verde, L., & Jimenez, R. 2009, Jour. Cos. and Part. Phys., 1, 10
Kerscher, M., Schmalzing, J., & Buchert, T. 1996, Mapping, measuring and modelling the universe, ed. P. Coles, V. Martinez, & M. J. P. Borderia, 247–252
Kim, A., Deustua, S., Gabi, S., et al. 1995, in Bulletin of the American Astronomical Society, Vol. 27, American Astronomical Society Meeting Abstracts, 1292
Kneib, J.-P. & Natarajan, P. 2011, Astronomy & Astrophysics, 19, 47
Knox, L. & Page, L. 2000, Physical Review Letters, 85, 1366
Koenderink, J. J. 1984, Biol. Cybern., 50, 363
Kofman, L., Pogosian, D., & Shandarin, S. 1990, Mon. Not. Royal Astro. Soc., 242, 200
Kofman, L. A. & Shandarin, S. F. 1988, Nature, 334, 129
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, The Astrophysical Journal Supplements, 180, 330
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, Astrophysical Journal Suppl., 192, 18
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2010, eprint arXiv:1001.4538
Komatsu, E. et al. 2011, Astrophysical Journal Supplement Series
Kreckel, K., Platen, E., et al. 2011, Astron.J., 141, 4
Kreckel, K., Platen, E., et al. 2012, Astron.J., 144, 16
Lam, T. Y., Desjacques, V., & Sheth, R. K. 2010, Mon. Not. Royal Astro. Soc., 402, 2397
Lavaux, G. & Wandelt, B. D. 2010, Mon. Not. Royal Astro. Soc., 403, 1392
Lavaux, G. & Wandelt, B. D. 2012, Astrophys. J. Lett., 754, 109
Lee, J. & Pen, U.-L. 2000, Astrophys. J. Lett., 532, L5
Lee, J. & Yi, S. 2015, in American Astronomical Society Meeting Abstracts, Vol. 225, American Astronomical Society Meeting Abstracts, #437.06
Lemaître, G. 1927, Annales de la Societe Scietifique de Bruxelles, 47, 49
Libeskind, N. I., Hoffman, Y., Knebe, A., et al. 2012, MNRAS, 421, L137
Limber, D. N. 1954, The Astrophysical Journal, 119, 655
Linde, A. D. 1982a, Physics Letters B, 114, 431
Linde, A. D. 1982b, Physics Letters B, 116, 335
Linder, E. V. & Jenkins, A. 2003, Mon. Not. Royal Astro. Soc., 346, 573
Lindner, U., Einasto, J., Einasto, M., et al. 1995, Astron. & Astrophysics, 301, 329
Lombardi, M. 2002, Astron. & Astrophysics, 395, 733
Lombardi, M. & Schneider, P. 2001, Astron. & Astrophysics, 373, 359

Lombardi, M. & Schneider, P. 2003, Astron. & Astrophysics, 407, 385
Martinez, V. J. 1990, Vistas in Astronomy, 33, 337
Martinez, V. J. & Coles, P. 1994, The Astrophysical Journal, 437, 550
Martinez, V. J., Jones, B. J. T., Dominguez-Tenreiro, R., & van de Weygaert, R. 1990, The Astrophysical Journal, 357, 50
Martínez, V. J. & Saar, E. 2002, Statistics of the Galaxy Distribution (Chapman)
Matsubara, T. 2010, Physical Review D, 81, 083505
Maximo, A., Marroquim, R., & Farias, R. 2010, in Proceedings of the 12th Eurographics / IEEE - VGTC conference on Visualization (Eurographics Association), 903–912
McDonald, P. 2008, Phys. Rev. D, 78, 123519
Mecke, K. R., Buchert, T., & Wagner, H. 1994, Astronomy & Astrophysics, 288, 697
Milnor, J. 1963, Journal of Mathematical Physics
Monaghan, J. J. 2005, Reports on Progress in Physics, 68, 1703
Moore, B., Frenk, C. S., Weinberg, D. H., et al. 1992, MNRAS, 256, 477
Moreland, K. & Angel, E. 2004, in Proc. of IEEE Symp. on Vol. Vis. and Graphics, 9–16
Mowbray, A. G. 1938, Publ. Astron. Soc. Pacific, 50, 275
Munkres, J. 1984, Elements of Algebraic Topology, Advanced book classics (Perseus Books)
Nagashima, M. & Gouda, N. 1997, Mon. Not. Royal Astro. Soc., 287, 515
Neben, A. R. & Turner, M. S. 2013, The Astrophysical Journal, 769, 133
Neumann, D. M., Lumb, D. H., Pratt, G. W., & Briel, U. G. 2003, Astron. & Astrophysics, 400, 811
Neyrinck, M. C. 2008, Mon. Not. Royal Astro. Soc., 386, 2101
Neyrinck, M. C. 2012, MNRAS, 427, 494
Oh, S. P., Spergel, D. N., & Hinshaw, G. 1999, ApJ, 510, 551
O'Hagan, A. & Leonard, T. 1976, Biometrika, 63, pp. 201
Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. 2000, Spatial tessellations: Concepts and applications of Voronoi diagrams, 2nd edn., Probability and Statistics (NYC: Wiley), 671 pages.
Park, C., Choi, Y.-Y., Vogeley, M. S., et al. 2005a, Astrophysical Journal, 633, 11
Park, C., Colley, W. N., Gott, III, J. R., et al. 1998, Astrophysical Journal, 506, 473
Park, C., Gott, III, J. R., & Choi, Y. J. 2001, Astrophysical Journal, 553, 33
Park, C., Kim, J., & Gott, III, J. R. 2005b, Astrophysical Journal, 633, 1
Park, C. & Kim, Y.-R. 2010, ApJ Letters, 715, L185
Park, C., Pranav, P., Chingangbam, P., et al. 2013, Journal of Korean Astronomical Society, 46, 125
Park, D. & Lee, J. 2007a, Physical Review Letters, 98
Park, D. & Lee, J. 2007b, The Astrophysical Journal, 665, 96
Parkinson, H., Cole, S., & Helly, J. 2008, Mon. Not. Royal Astro. Soc., 383, 557
Peacock, J. A. & Heavens, A. F. 1990, Mon. Not. Royal Astro. Soc., 243, 133
Peebles, P. 1980, The Large-scale Structure of the Universe, Princeton series in physics (Princeton University Press)
Peebles, P. J. E. 1969, Astrophys. J., 155, 393

Peebles, P. J. E. 1993, Principles of Physical Cosmology
Peebles, P. J. E. 2001, Astrophys. J, 557, 495
Peebles, P. J. E. & Ratra, B. 1988, The Astrophysical Journal, 325, L17
Peebles, P. J. E. & Yu, J. T. 1970, ApJ, 162, 815
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, The Astrophysical Journal, 517, 565
Peterson, J. B., Carlstrom, J. E., Cheng, E. S., et al. 1999, ArXiv Astrophysics e-prints
Phillips, J. M., Wang, B., & Zheng, Y. 2013, ArXiv e-prints
Pisani, A., Sutter, P. M., Hamaus, N., et al. 2015, ArXiv e-prints
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013a, ArXiv e-prints
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013b, ArXiv e-prints
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, ArXiv e-prints
Platen, E., van de Weygaert, R., & Jones, B. J. T. 2007, MNRAS, 380, 551
Pogosyan, D., Gay, C., & Pichon, C. 2009, Phys. Rev. D, 80, 081301
Pranav, P., Edelsbrunner, H., van de Weygaert, R., & et al. 2013, MNRAS, to be subm.
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Protogeros, Z. A. M. & Weinberg, D. H. 1997, Astrophysical Journal, 489, 457
Ramaley, J. F. 1969, The American Mathematical Monthly, 76, 916
Reininghaus, J., Kotava, N., et al. 2011, IEEE Trans. Vis. & Comp. Graphics, 17, 2045
Rieder, S., van de Weygaert, R., Cautun, M., Beygu, B., & Portegies Zwart, S. 2013, MNRAS, 435, 222
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, The Astronomical Journal, 116, 1009
Ripley, B. D. 1981, Spatial statistics
Robins, V., Wood, P., & Sheppard, A. 2011, IEEE Trans. on Pattern Analysis and Machine Intelligence, 33, 1646
Rubin, V. C. 1954, Proceedings of the National Academy of Science, 40, 541
Ryden, B. S. & Melott, A. L. 1996, Astrophys. J., 470, 160
Sahni, V., Sathyaprakash, B. S., & Shandarin, S. F. 1997, The Astrophysical Journal, 476, L1
Sahni, V., Sathyprakash, B., & Shandarin, S. 1998, Astrophys. J., 507, L109
Saslaw, W. C., ed. 2000, The distribution of the galaxies : gravitational clustering in cosmology
Scaramella, R. & Vittorio, N. 1991, Astrophysical Journal, 375, 439
Schaap, W. E. & van de Weygaert, R. 2000a, arXiv, 363, L29
Schaap, W. E. & van de Weygaert, R. 2000b, Astronomy and Astrophysics, 363, L29
Schaap, W. E. & van de Weygaert, R. 2001, ArXiv Astrophysics e-prints
Schmalzing, J. & Buchert, T. 1997, Astrophys. J. Lett., 482, L1+
Schmalzing, J., Buchert, T., Melott, A., et al. 1999, Astrophys. J., 526, 568
Schmidt, J. D., Ryden, B. S., & Melott, A. L. 2001, Astrophys. J., 546, 609
Seljak, U., Mandelbaum, R., & McDonald, P. 2002, ArXiv Astrophysics e-prints
Seljak, U. & Zaldarriaga, M. 1999a, CMBFAST: A microwave anisotropy code, Astrophysics Source Code Library
Seljak, U. & Zaldarriaga, M. 1999b, Physical Review Letters, 82, 2636
Shandarin, S., Habib, S., & Heitmann, K. 2012, Phys. Rev. D, 85

Shandarin, S. F. 2011, JCAP, 5, 15
Shandarin, S. F. & Zeldovich, I. B. 1983, Comments on Astrophysics, 10, 33
Sheth, R. K. 1998, Mon. Not. Royal Astro. Soc., 300, 1057
Sheth, R. K. & Tormen, G. 2002, Mon. Not. Royal Astro. Soc., 329, 61
Sheth, R. K. & van de Weygaert, R. 2004, MNRAS, 350, 517
Shimizu, M., Kitayama, T., Sasaki, S., & Suto, Y. 2002, Publ. Astron. Soc. Japan, 54, 645
Shivashankar, N. & Natarajan, V. 2012, Computer Graphics Forum, 31, 965
Shivashankar, N., Senthilnathan, M., & Natarajan, V. 2012, IEEE Trans. Vis. & Comp. Graphics, 18, 1757
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, The Astronomical Journal, 131, 1163
Smoot, G. F., Bennett, C. L., Kogut, A., & et. al. 1992, Astrophys.J.Lett., 396, L1
Soneira, R. M. & Peebles, P. J. E. 1978, The Astronomical Journal, 83, 845
Song, H. & Lee, J. 2009, The Astrophysical Journal, 701, L25
Sousbie, T. 2011, MNRAS, 414, 350
Sousbie, T., Pichon, C., & Kawahara, H. 2011, MNRAS, 414, 384
Spergel, D. N., Bean, R., Doré, O., & et. al. 2007, Astrophys.J.Suppl., 170, 377
Springel, V. 2005, MNRAS, 364, 1105
Springel, V. 2008, in The Eleventh Marcel Grossmann Meeting On Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories, ed. H. Kleinert, R. T. Jantzen, & R. Ruffini, 309–325
Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Stoica, R., Gregori, P., & Mateu, J. 2005, Stochastic Processes and their Applications
Stoica, R. S., Martínez, V. J., Mateu, J., & Saar, E. 2005, Astron. & Astrophysics, 434, 423
Stoica, R. S., Martínez, V. J., & Saar, E. 2007, Journal of the Royal Statistical Society: Series C (Applied Statistics) 56 (4), 459-477, 56
Sutter, P. M., Lavaux, G., Wandelt, B. D., et al. 2014a, Mon. Not. Royal Astro. Soc., 442, 3127
Sutter, P. M., Pisani, A., Wandelt, B. D., & Weinberg, D. H. 2014b, MNRAS, 443, 2983
Szapudi, I. & Szalay, A. S. 1993a, ApJ, 414, 493
Szapudi, I. & Szalay, A. S. 1993b, The Astrophysical Journal, 408, 43
Szapudi, I. & Szalay, A. S. 1997, The Astrophysical Journal, 481, L1
Taylor, J. E. & Adler, R. J. 2009, The Annals of Probability, 37, 1459
Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, Physica Rev. D., 69, 103501
Tempel, E. & Libeskind, N. I. 2013, The Astrophysical Journal, 775, L42
Tempel, E., Libeskind, N. I., Hoffman, Y., Liivamägi, L. J., & Tamm, A. 2014, Mon. Not. Royal Astro. Soc., 437, L11
Tempel, E., Stoica, R. S., & Saar, E. 2013, Mon. Not. Royal Astro. Soc., 428, 1827
Terukina, A. & Yamamoto, K. 2012, Phys. Rev. D, 86, 103503
Tomita, H. 1993, in Formation, Dynamics, and Statistics of Patterns, ed. K. Kawasaki & M. Suzuki, Vol. 1, 113
Totsuji, H. & Kihara, T. 1969, Publ. Astron. Soc. Japan, 21, 221

Tristram, M., Patanchon, G., Macías-Pérez, J. F., & et. al. 2005, Astronomy and Astrophysics, 436, 785
Tyson, J. 2000, in Encyclopedia of Astronomy and Astrophysics, ed. P. Murdin
van de Weygaert, R. 1991, Voids and the geometry of large scale structure (PhD thesis, University of Leiden)
van de Weygaert, R. 1994, Astron. & Astrophysics, 283, 361
van de Weygaert, R. 2002, in Astrophysics and Space Science Library, Vol. 276, Modern Theoretical and Observational Cosmology, ed. M. Plionis & S. Cotsakis, 119
van de Weygaert, R. & Bond, J. R. 2008, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 740, A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure, ed. M. Plionis, O. López-Cruz, & D. Hughes, 335
van de Weygaert, R. & Icke, V. 1989, Astron. Astrophys., 213, 1
van de Weygaert, R., Kreckel, K., et al. 2011, in Environment and the Formation of Galaxies: 30 Years Later, ed. I. Ferreras & A. Pasquali, 17
van de Weygaert, R. & Platen, E. 2011, International Journal of Modern Physics Conference Series, 1, 41
van de Weygaert, R. & Schaap, W. 2009a, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 665, Data Analysis in Cosmology, ed. V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería, 291–413
van de Weygaert, R. & Schaap, W. 2009b, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 665, Data Analysis in Cosmology, ed. V. J. Martínez, E. Saar, E. Martínez-González, & M.-J. Pons-Bordería, 291–413
van de Weygaert, R. & van Kampen, E. 1993, MNRAS, 263
van de Weygaert, R., Vegter, G., Edelsbrunner, H., et al. 2011, Transactions on Computational Science, 14, 60
Verde, L., Wang, L., Heavens, A. F., & Kamionkowski, M. 2000, Mon. Not. Royal Astro. Soc., 313, 141
Vogeley, M. S., Park, C., Geller, M. J., Huchra, J. P., & Gott, III, J. R. 1994, Astrophysical Journal, 420, 525
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, ArXiv e-prints
von Hoerner, S. 1960, Z. Astrophys., 50, 184
Vorono"i, G. 1908, Journal für die reine und angwandte Mathematik, 133, 97
Weinberg, D. H. 1988, PASP, 100, 1373
Weinberg, D. H., Gott, III, J. R., & Melott, A. L. 1987, Astrophysical Journal, 321, 2
Weinkauf, T. & Günther, D. 2009, in Proc. of Symp. on Geom. Proc., 1519–1528
Weygaert, R. v. d. 2007, in ISVD '07: Proc. of Symp. on Voronoi Diagrams in Science and Engineering (Washington, DC, USA: IEEE Computer Society), 230–239
White, S. D. M. 1979, Mon. Not. Royal Astro. Soc., 186, 145
White, S. D. M. 1984, Astrophys. J., 286, 38
White, S. D. M., Efstathiou, G., & Frenk, C. S. 1993, MNRAS, 262, 1023
White, S. D. M. & Silk, J. 1979, The Astrophysical Journal, 231, 1
Wilson, G. W., Knox, L., Dodelson, S., et al. 2000, ApJ, 532, 57
Wintraecken, M. & Vegter, G. 2013, Topology and its Applications Subm.
Wintraecken, M., Vegter, G., van de Weygaert, R., & Jones, B. J. T. 2013, Journal of Mathematical Physics subm.

Wintraecken, M. H. M. J. 2012, Betti numbers cannot be expressed as Euler integrals
Zel'dovich, Y. B. 1970, Astronomy and astrophysics, 5, 84
Zeldovich, Y. B. 1972, MNRAS, 160, 1P
Zomorodian, A. & Carlsson, G. 2005, Discrete & Computational Geometry, 33, 249

Summary

1 The Cosmic web

While the cosmic matter distribution is uniform on large Gigaparsec scales, it has an intricate web-like pattern on smaller distances. At scales of few to hundreds of megaparsecs, [1] the universe has a web-like appearance, commonly known as the *Cosmic Web*. That the cosmic web indeed exists has been confirmed by a host of galaxy redshift surveys like the CfA and CfA2 surveys (de Lapparent et al. 1986; Huchra & Geller 1982), the 2dFGRS survey (Colless et al. 2003), the SDSS survey (Abazajian et al. 2003) and the 2MRS survey (Huchra et al. 2005; Skrutskie et al. 2006). Massive megaparsec to gigaparsec scale computer simulations also show the existence of the Cosmic web (Springel 2008). Figure 1 presents an impression of the Cosmic Web at a range of scales, as revealed in the state-of-the-art millennium simulation.

The cosmic web is composed of huge, virialized blobs called *clusters*, containing thousands of galaxies, as well as huge amounts of dark matter and high temperature gas. Out of these virialized clusters, spread out gigantic *filaments*. These filaments, also containing thousands of galaxies, dark matter and gas, serve as transport channels for mass flow in the Universe. The cosmic web also consists of huge sheet-like objects commonly known as *walls*. The clusters, filaments and walls of the web are intertwined around vast near-empty *cosmic voids*. Figure 2 presents a view of the observational reality of the large scale structure of the Universe at 0.45¡z¡1.1, provided by the distribution of 25,000 galaxies as revealed by the VIPERS survey (Guzzo & The Vipers Team 2013). The presence of cluster-like and filamentary patterns is clearly visible.

According to the standard paradigm of structure formation, the *gravitational instability scenario*, cosmic structure grows from tiny primordial density and velocity

[1] A parsec is the standard unit of measurement of distances in the cosmos. A parsec is 3.26 times the *light-year*, the distance light covers in a year. A megaparsec is a million parsecs, the typical scale of measurement of size of the large scale structures in the universe.

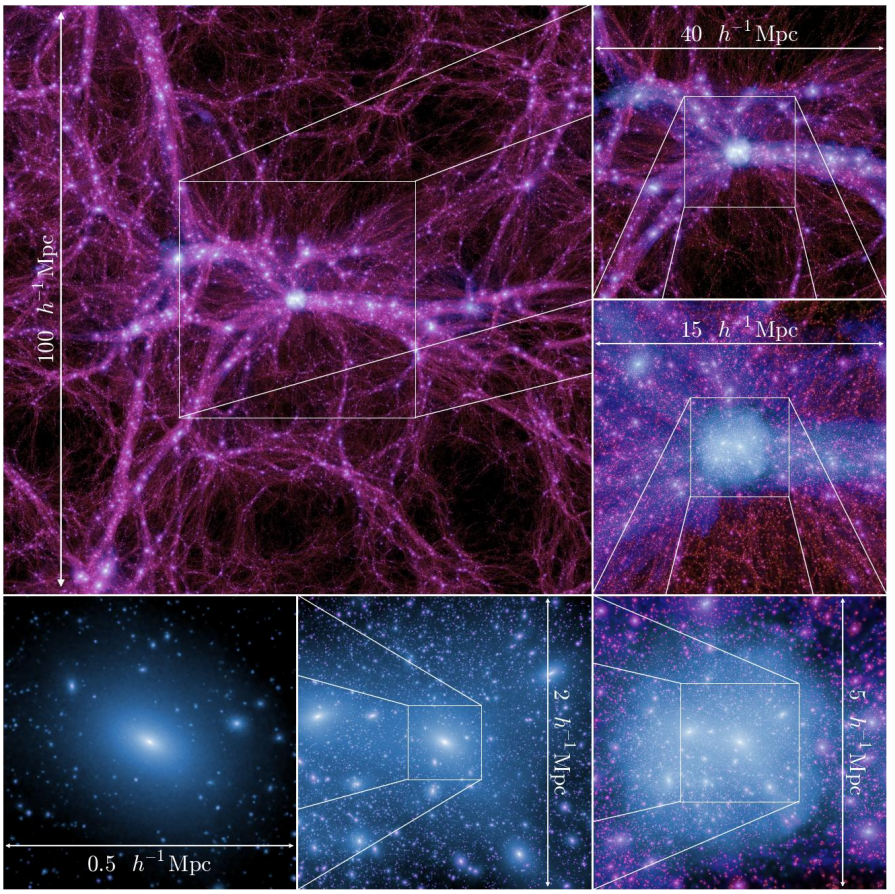


Figure 1 A sequential zoom through the Millennium-II Simulation depicting the hierarchical nature of structures in the Universe. This is a dark matter simulation revealing the anisotropic and multiscale character of the hierarchically evolving mass distribution. The number of particles is over 10 billion approximately. The large image (upper left) is a $15h^{-1}$Mpc thick slice through the full $100h^{-1}$Mpc simulation box at redshift zero, centered on the most massive halo in the simulation. The clusters, filaments, walls and the massive voids are clearly visible. The successive frames zoom in to smaller regions. A rich variety of substructures is visible at all scales. Figure courtesy Boylan-Kolchin et al. (2009).

perturbations. The evidence provided by the temperature fluctuations in the cosmic microwave background (Planck Collaboration et al. 2013b) suggests that the character of the perturbation field is that of a homogeneous and isotropic spatial Gaussian process. Theories of the early universe predict such primordial Gaussian perturbations as a natural product of an early inflationary phase of our Universe.

The vast Megaparsec Cosmic Web is one of the most striking examples of complex geometric patterns found in nature, and certainly the largest in terms of size. The features of the web are highly anisotropic in nature. They are also characterized by a wide variety of shapes and dimensions. The clusters are virialized near-

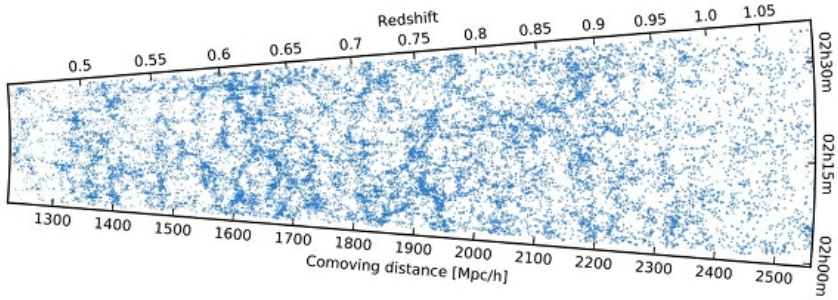


Figure 2 A view of the observational reality of the large scale structure of the Universe at $0.45 < z < 1.1$, provided by the distribution of $\sim 25,000$ galaxies as revealed by the VIPERS survey. The survey clearly indicates the presence of cluster-like and filamentary structures. Figure courtesy Guzzo & The Vipers Team (2013).

spherical objects, that represent the most evolved stages of the matter distribution. The filaments and walls lack symmetry, with typically one or two dimensions much larger than the remaining. The most prominent morphological element of the web in terms of the volume occupancy is the near-empty vast cosmic voids. The morphological elements of the web also exhibit an asymmetry with respect to the mass fraction and volume occupancy. The clusters contain the highest mass fraction, while occupying the lowest volume. On the other hand, the voids contain the lowest mas fraction, but have the highest volume occupancy. This is in contradiction to the Gaussian primordial conditions which are symmetric in the fractional volume occupancy of the underdense and overdense regions.

All the elements of the web are ubiquitous at all ranges of scales and densities. This is the clearest manifestation of the hierarchical nature of the evolution of the matter distribution in the cosmos. Finally, the web is marked by distinctly complex connectivity characteristics. The low dimensional morphological elements of the web form the boundaries of the higher dimensional elements, resulting in a pervasive network of complex and inter-connected patterns.

It has remained a major challenge to characterize the structure, geometry, and topology of the Cosmic Web. The overwhelming complexity of the individual structures as well as their connectivity, the lack of structural symmetries, the intrinsic multi-scale nature of the web, and the wide range of densities in the cosmic matter distribution has prevented the use of simple and straightforward instruments. Below we present a brief but non-exhaustive summary of the methods of analysis of the web.

2 Characterization of the cosmic web

The effort to describe and characterize the mass distribution in the Universe has formed a decades long challenge. A quantitative characterization of the complex and rich spatial patterns in the Universe is extremely important in view of comparing

and validating various models with respect to the actual mass distribution in the Universe.

The 2-point correlation function or the autocorrelation function, has been the mainstay of statistical analysis in cosmology (Peebles 1980). For an excellent and comprehensive textbook on the statistical characterization of the large scale mass distribution, see Martínez & Saar (2002). However, the autocorrelation function is not sensitive to the phases of the Fourier modes. As a result, it is possible for two distributions with widely varying geometric and morphological patterns to have the same correlation function. Higher order correlation functions have also been used to describe the statistics of matter distribution. However, they are limited in their usefulness because the errors in their measurement become prohibitively large with increasing order.

Over the last decade the view has emerged that a full analysis of the rich structures in the Universe need to take into account the geometric and morphological aspects of the elements of the Cosmic Web. With this view, there has been an effort to detect and isolate the morphological components of the web, to study their characteristics and connectivity in a more systematic way.

Within this, the statistical measures include the analysis of the global connectivity of Cosmic Web, by means of *percolation analysis* (Shandarin & Zeldovich 1983) and the *minimal spanning tree* (Barrow et al. 1985) of the galaxy distribution. Percolation analysis and spanning trees, while relatively sensitive to the underlying structure, are still global quantities and fail to capture and describe the local variations in shape.

Noteworthy also are the models and methodologies emanating from statistical and geometric considerations for delineating the cosmic filaments. Most prominent among them are the *candy model*, that uses a marked point process to delineate the filaments (Stoica et al. 2005; Tempel et al. 2013), and the recovery of filaments by recognizing them as the medial axis of a given point process (Genovese et al. 2010).

Morphological methods aimed at determining shapes locally considering the properties of the eigenvalues of the Hessian of a range of fields arising in cosmology have also been at the forefront in delineating the cosmic web. These involve the eigenvalues of the Hessian of density field (Aragón-Calvo et al. 2007b; Cautun et al. 2013) or those of the tidal tensor or the velocity shear tensor (Tempel et al. 2014). A variation of these methods further evaluates this in a multi-scale fashion by performing the procedure on a stack of images in the scale space, which are derived from the original image through convolution with an appropriate filter at a range of scales (Aragón-Calvo et al. 2010; Cautun et al. 2014).

Morphology based shape determination methods are closely related to the shape determination methods arising from topological considerations. Both the methods involve the identification of a morphological element by considering the properties of the eigenvalues of the Hessian of the density field for either extended objects (morphological methods), or that of a localized point in space (topological methods).

Besides being useful for delineating the structures in the web, topological methods are also important from the point of view of investigating the inherent structure and connectivity between the various structural elements that form the complex pat-

terns that is the Cosmic Web. In the next section, we present a brief account of topology, with a focus on cosmological applications. This is the main theme of this thesis.

3 Topology of the mass distribution in the Universe

Topology is the branch of mathematics that is concerned with the properties of space that are preserved under continuous deformations including stretching (compression) and bending, but not tearing or gluing. It also includes invariance of properties such as connectedness and boundary.

This thesis concerns itself with a topological analysis of scalar fields that model various aspects of the cosmic mass distribution. The strategy is to define such functions or fields on a manifold, and study the topology that this function induces on the manifold. In the context of this thesis, the manifold is either $\mathbb{R}^3$ or $\mathbb{T}^3$. The latter is formed by identifying and glueing opposite faces of a cube.

We study the topology of the manifold by tracking the topological changes for a growing superlevel set [2]. The critical points of the scalar field are responsible for the topological changes induced in the manifold[3] (Milnor 1963). The nature and distribution of critical points can reveal insightful information about the nature of the field itself. As an example, a symmetric distribution of critical points about the mean of the density field suggests the presence of a Gaussian field.

In the context of the Cosmic Web, it is known that the individual strutural elements of different dimensions connect together in a complex and hierarchical fashion. Topology has the ability to characterize shapes of different dimensions. We can also describe the complex structural elements, and how they connect to form the web, through topological concepts. The analysis of the connectivity characteristics help to develop a better understanding of how the large scale structures of the Cosmic Web emerge as neighboring structures of smaller scale interact. Mergers and possibly fragmentations are the most common instances of such interactions.

Because topology is insensitive to continuous deformations like stretching and bending, topological measures are also expected to be relatively insensitive to systematic effects such as non-linear gravitational evolution, galaxy biasing, and redshift-space distortion (Park & Kim 2010). As a result, one expects the topology of the primordial field and the large-scale linear field at the later epochs to be similar (Gott et al. 1986).

Topological studies have been invoked to characterize the cosmic mass distribution for many decades. The early studies focused on topological characterization of the cosmic mass distribution by evaluating the genus and the Euler characteristic of the iso-density surfaces, as a function of the density threshold (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986). Later, more discriminative information became available with the introduction of Minkowski functionals (Mecke et al. 1994). It is important to note that, these studies have largely been of a heuristic nature with a primary focus on model discrimination. In the fol-

[2]For a given function value ν, the *superlevel set* or *excursion region* is the part of the manifold that is above the threshold ν. The boundary of this region is called the *level set* or *excursion set*

[3]A critical point of a function is the location where the gradient of the function vanishes, i.e. $\nabla f = 0$.

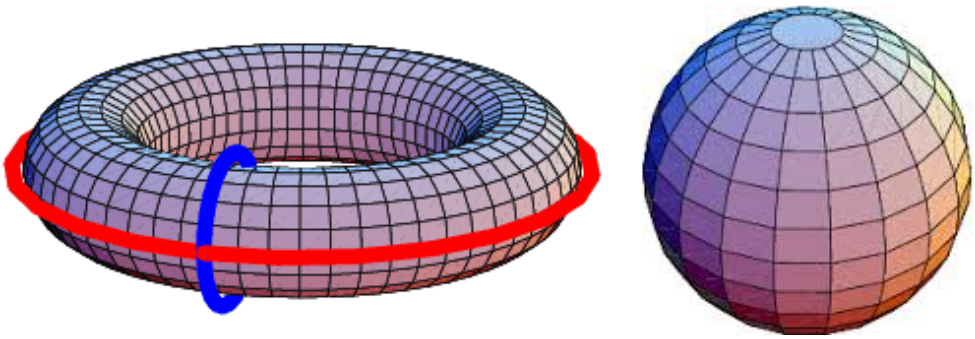


Figure 3 Illustration of the concept of Genus. Genus is the number of independent closed curves that can be drawn on a surface, without rendering it disconnected. For a torus, there can be two independent closed curves that can be drawn without rendering it disconnected (drawn in red and blue). This suggests that the genus of a torus is 2. It is not possible to draw any closed curve on the surface of a sphere without rendering it disconnected. Hence, the genus of a sphere is 0.

lowing we elaborate on the key topological measures genus, Euler and Minkowski functionals.

3.1 Genus

For a connected, orientable surface, the *Genus* [4] is defined as the maximal number of independent simple closed curves that can be drawn on the surface without rendering it disconnected.

We give an intuitive interpretation of genus through the illustration of a torus and a sphere in Figure 3. There can be two independent closed curves along which the surface of a torus can be cut without rendering it disconnected. These curves are drawn in red and blue. This suggests that the genus of a torus is 2. On the other hand, it is not possible to draw any closed curve on the surface of a sphere without rendering it disconnected. Hence, the genus of a sphere is 0.

3.2 Euler characteristic

The genus is intimately related to perhaps the most fundamental quantity in topology – the *Euler characteristic*. Let a closed surface $\mathbb{M}$ have genus g. Then the *Euler formula* states that

$$\chi(\mathbb{M}) = V - E + F, \tag{5.8}$$

where $\chi(\mathbb{M})$ is the Euler characteristic, sometimes also known as the *Euler-Poincaré characteristic*. The above equation is for polyhedral surfaces in 3D space. V, E and F are the number of vertices, edges and faces used to triangulate the manifold. In a more general language, the Euler characteristic of a manifold is the total number

[4] For consistency, it is important to note that the definition of genus g used in cosmological studies is different from the mathematical definition of genus G, the difference being the number of connected regions: $g = G - c$.

of even dimensional simplices minus the total number of odd dimensional simplices needed to triangulate it. The Euler characteristic originated from the analysis of polyhedral objects.

As an example, let us consider a sphere. Topologically, this is equivalent to a tetrahedron. The tetrahedron has 4 vertices, 6 edges, and 4 faces or triangles. So, the Euler characteristic of a sphere (or a tetrahedron) is 2. In fact, any convex polygon, as e.g., a cube, octahedron or a dodecahedron will have the same Euler characteristic as a tetrahedron or a sphere. Topologically, they are the same object.

3.3 Minkowski functionals

Minkowski functionals relate to the geometric properties of a manifold. Suppose we have a solid body, $\mathbb{M}$, whose boundary is a smoothly embedded surface in $\mathbb{R}^3$. This surface may be a sphere or have holes, like the torus, and it may consist of one or several connected components, each with its own holes. Similarly, we do not require that $\mathbb{M}$ is connected. Write $\mathbb{M}^r$ for the set of points at distance r or less from $\mathbb{M}$. For small values of r, the boundary of $\mathbb{M}^r$ will be smoothly embedded in $\mathbb{R}^3$, but as r grows, it will develop singularities and self-intersections. Before this happens, the volume of $\mathbb{M}^r$ can be written as a degree-3 polynomial in r,

$$\text{vol}\,\mathbb{M}^r \;=\; Q_0 + Q_1 r + Q_2 r^2 + Q_3 r^3\,. \tag{5.9}$$

The Q_i are known as the *Minkowski functionals* of $\mathbb{M}$, which are important concepts in integral geometry. For a d-dimensional manifold $\mathbb{M}$ there are $(d+1)$ Minkowski functionals, $Q_k((k = 0, \ldots, d)$ (Munkres 1984). In the context of this thesis, we are interested in $d \leq 3$. In 3D, Q_0 is the volume of $\mathbb{M}$, Q_1 is the area of its boundary, Q_2 is the total mean curvature, and Q_3 is one third of the total Gaussian curvature of the boundary. According to the Gauss-Bonnet theorem, Q_3 is equal to the Euler characteristic of the boundary of $\mathbb{M}$ (up to a multiplicative constant).

3.4 Beyond the Euler characteristic and Minkowski functionals

Genus and Euler characteristic, along with the rest of the Minkowski functionals, have provided crucial insights into the topology and morphology of cosmic density fields from models as well as observational data. For example, the Euler characteristic of the cosmic density fields in the linear regime suggest the presence of near-Gaussian conditions in the early Universe. Nonetheless, there are a range of grounds that make it necessary to explore a more elaborate and extensive arsenal of topological concepts and descriptions.

First, we note that the structures in the Universe form and evolve in a hierarchical fashion (Peebles 1980). Smaller high density structures coalesce together hierarchically to build up larger structures of lower density. Genus, Euler characteristic and Minkowski functionals are not equipped to address this aspect. A formalism capable of adressing topology in an intrinsically hierarchical context would present an interesting and powerful extension to the arsenal of tools used for describing the hierarchical buildup of structures in the cosmos.

Second, the Euler-Poincaré formula states that the Euler characteristic is an alternating sum of *Betti numbers* (Section 5.1.1). This indicates the Betti numbers in

essence have a richer information content than the Euler characteristic and genus.

Third, while the genus can distinguish between connected, closed surfaces in $\mathbb{R}^3$, it has no discriminative power if applied to a 3-manifold. In other words, the genus is defined only for connected and closed 2-dimensional surfaces, and has no generalizations in higher or lower dimensions. This is not desirable in the context of cosmic density fields where we extensively deal with 3D manifolds.

4 Homology, Morse theory and Persistence

The extension of our topological arsenal concerns the formalisms of *homology*, *Morse theory* and *persistence* in the cosmological context. In this thesis, we demonstrate that they have the potential of uncovering a more wealthy and detailed topological information content about the cosmic mass distribution.

4.1 Homology and Betti numbers

Homology theory is a mathematical formalism associating computable algebraic objects (groups) to topological spaces [5], through assessing the boundaries of a manifold (Munkres 1984). Homology groups provide a mathematical language for the holes in a topological space (Edelsbrunner & Harer 2010). A d-manifold can be composed of topological holes of 0 up to $(d-1)$ dimensions. Holes in $d < 3$ have intuitive interpretations. A 0-dimensional hole is a *gap* between two isolated independent objects. A 1-dimensional hole is a *tunnel* through which one can pass in any one direction without encountering a boundary. A 2-dimensional hole is a *cavity* or *void* fully enclosed within a 2-dimensional surface. This intuitive interpretation in terms of 'gaps' and 'tunnels' is only valid for surfaces embedded in $\mathbb{R}^3$, $\mathbb{S}^3$ or $\mathbb{T}^3$.

Homology can also be defined in terms of the *cycles* that form the boundary of a hole. In 3D, A 0-cycle is a connected object (and hence, the boundary of a 0-hole, which is the gap between these two independent objects). A 1-cycle is a loop/cylinder that surrounds a tunnel. A 2-cycle is a *shell* enclosing a void.

The collection of all p-dimensional cycles of the manifold is the p-th homology group $\mathbb{H}_p$. The rank of this group is the collection of all *independent* cycles, denoted by the *Betti numbers* β_p, where $p = 0, \ldots, d$ (Betti 1871; Edelsbrunner & Harer 2010). In 3D, β_0 counts the number of independent components, β_1 counts the number of loops enclosing the independent tunnels and β_2 counts the number of shells enclosing the independent voids.

The Betti numbers have extra information content about topology than the genus or the Euler characteristic. The Euler-Poincaré Formula states that the Euler characteristic can be written as an alternating sum of the Betti numbers (Edelsbrunner & Harer 2010; Adler & Taylor 2010):

$$\chi \;=\; \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{5.10}$$

where d is the intrinsic dimension of the manifold.

[5] There is a notion of k-connectedness, $k = 0, \ldots, d$, where d is the dimension of the manifold. Within this, 0-connectedness is the same as the 'usual' notion of connectedness.

4.2 Morse theory

Morse theory relates the properties of a scalar function f to the topological changes it induces on a manifold on which it is defined (Milnor 1963). The critical points of the function are responsible for the topological changes. Morse theory has deep connections with the distribution of critical points and the ensuing singularity structure of the function. The *critical points* are the location where the gradient of f vanishes, i.e., $\nabla f = 0$. The function f is said to be a *Morse function* if all of its critical points are non-degenerate i.e., the *Hessian* of f, defined as the matrix of second order partial derivatives, is non-singular.

Using a coordinate transformation, the function near a critical point can be written as a quadratic function without mixed terms. The *index* of a critical point is equal to the number of negative terms in this expression. The nature of the critical point defines the geometry of the manifold in its neighborhood. In 2D, the index 0 corresponds to a minimum, the index 1 corresponds to a saddle point and index 2 corresponds to a maximum. In 3D, the index 0 corresponds to minima, the index 1 corresponds to 1-saddles, the index 2 corresponds to 2-saddles, and the index 3 corresponds to maxima. Figure 4 illustrate the geometry and the behavior of the gradients of the manifolds in the neighborhood of various kinds of critical points in 3D.

A major theorem of Morse theory establishes that the global shape and topology of a manifold is defined only by it critical points (Milnor 1963; Edelsbrunner & Harer 2010). All changes in topology of the superlevel set of $\mathbb{M}$ occur only when the corresponding level set passes through a critical point of f. It also implies that topological changes occur in discrete steps.

4.3 Persistence

Stemming from homology and Morse theory, the formalism of *persistence* is inherently hierarchical. The hierarchical expression of topology in persistence is achieved by constructing a *filtration* of the cosmic density fields. This can be done by continuously decreasing the superlevel set. For a given manifold $\mathbb{M}$, and a density threshold ν, the superlevel set $\mathbb{M}^s(\nu)$ is the subset of $\mathbb{M}$ with density $\nu_s \geq \nu$. Decreasing the superlevel set entails that a manifold corresponding to a higher density threshold is necessarily embedded in a manifold corresponding to a lower density threshold.

As an example, consider the manifolds corresponding to two different density thresholds of a 2D scalar function in Figure 5. The manifold at the higher density threshold in panel (a) consists of three disjoint islands that grow around the local maxima of the function. In the manifold at the lower density threshold in panel (b), two of the three components merge. The process of merger occurs due to the introduction of a saddle point in the manifold, as the density threshold decreases. The manifold corresponding to the higher threshold is embedded in the manifold corresponding to the lower threshold. This is evident once we realize that the former can be recovered from the latter by removing the saddle point that was responsible for the merger.

The addition of a critical point with index p can result in either the birth of a p-dimensional hole or the death of a $(p-1)$-dimensional hole. Each hole is associated

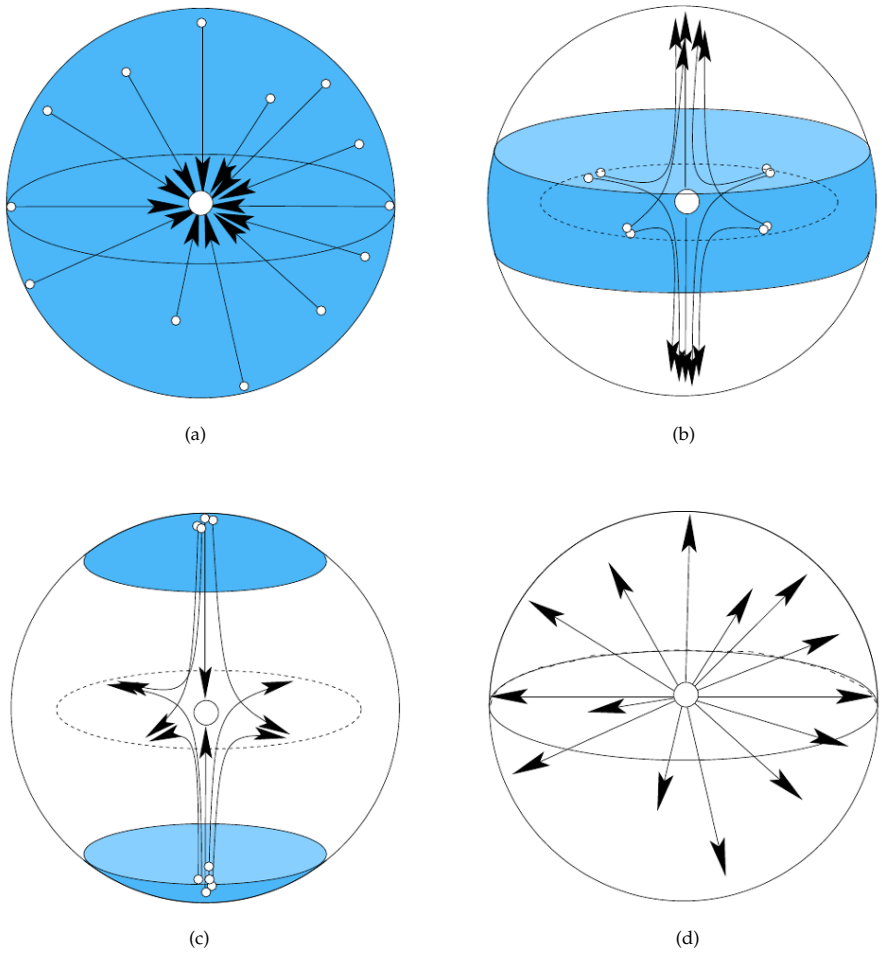
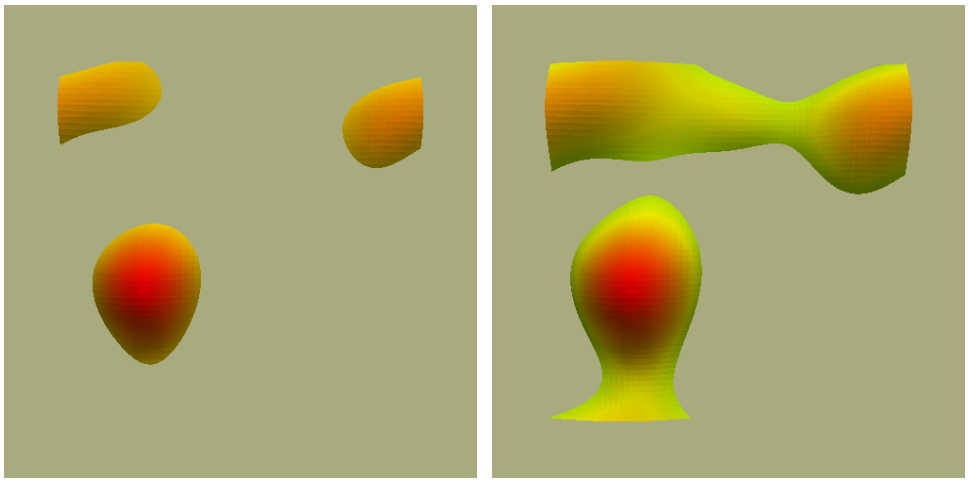


Figure 4 Critical points in 3D. Panels (a) through (d) present the maximum, 2-saddle, 1-saddle and the minimum. For a maximum, the flow is directed towards it. For a minimum, the flow is directed away from it. For a 1-saddle, the flow is towards it in a plane, and away from it along a linear element. The directions reverse for a 2-saddle.

with two unique function values: $f(c_b)$ associated with the critical point c_b that gives birth to it, and $f(c_d)$ associated with the critical point c_d that is responsible for *killing* or filling up the hole. The *life-span* of the hole, or its *persistence* is then defined as the absolute difference between the death and the birth values. In general, higher persistence holes may embed holes of lower persistence within themselves.

Persistence homology is represented in terms of *persistence diagrams* (Edelsbrunner & Harer 2010), which is a collection of dots in $\mathbb{R}^2$. Each dot is associated with a unique topological change in the manifold. There is a diagram for each ambient di-



(a) (b)

Figure 5 Birth and death for islands and tunnels. Panel (a) shows three independent islands. In panel (b), two of the three islands merge, signifying the death of a 0-dimensional hole. In the filtration, the manifold in panel (a) is embedded in the manifold in panel (b).

mension of the manifold. The dots in the 0-dimensional diagram record the merger events of two isolated objects. The dots in the 1-dimensional diagrams record the formation and destruction of loops, while the dots in the 2-dimensional diagrams record the birth and death of topological voids. Figure 6 presents a typical persistence arising from a density field arising from a stochastic process.

Persistence diagrams contain a wealth of information. The 0-dimensional diagram records merger events, that are a characteristic property of hierarchically evolving mass distributions. The 1-dimensional diagrams encode the topological signals of the percolation properties. The 2-dimensional diagrams encode the hierarchical

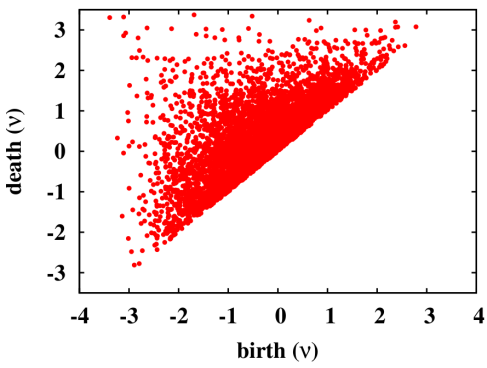


Figure 6 A typical persistence diagram arising from a realization of Gaussian field model. It is a scatter plot where each dot represents a unique topological hole. The coordinates of the dot correspond to the critical points responsible for the birth and death of the hole.

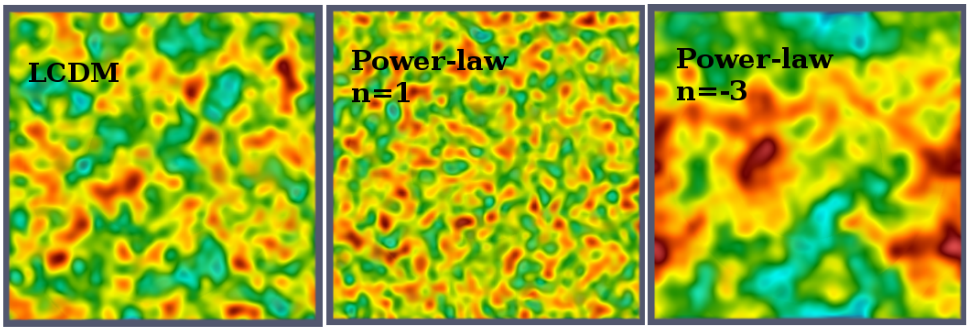


Figure 7 2D slices of a single realization of 3D Gaussian random field models. The models are constructed in a simulation box of side $128h^{-1}$Mpc with a grid resolution of $1h^{-1}$Mpc. Subsequently, they are smoothed with a Gaussian kernel of scale $R_f = 2h^{-1}$Mpc. The left panel shows a realization of the LCDM power spectrum. The the middle and the right panels panels show realizations of power-law power spectra with spectral indices $n = 1$ and $n = -3$.

aspects of formation and merger of voids. Additionally, this information is presented in a hierarchical fashion, where in small-scale voids with higher average density are the constituents of large scale voids with lower average density. This relates to the recent discovery that the cosmic voids are continuously evolving through mergers and other complex processes (Sheth & van de Weygaert 2004). However, an assessment is required to ascertain the exact correspondance between topological voids and cosmological voids.

5 Cosmic homology and persistence: results

This thesis has focused on the topological analysis of a few cosmologically relevant mass distribution models via homology, Morse theory and persistence. Below, we present a summary of the results.

5.1 Convergence of persistence diagrams of stochastic processes.

In Chapter 2 and Chapter 3 we present an epirical demonstration that the ensemble average of persistence diagrams arising from stochastic processes are convergent to a stable average. We demonstrate this by means of a few representative examples of stochastic processes, namely the Poisson point process, Gaussian random fields, the Voronoi models and the Soneira-Peebless model (described later in the section).

5.2 Homology and persistence of Gaussian fields

Gaussian fields are important in the description of the primordial fluctuation field. A Gaussian field is fully specified by its power spectrum. In this thesis, we analyze the models specified by a power-law power spectrum as well as the LCDM spectrum. Figure 7 presents 2D slices of a single realization of a 3D Gaussian field specified by the different power spectra. As the index of the power spectrum de-

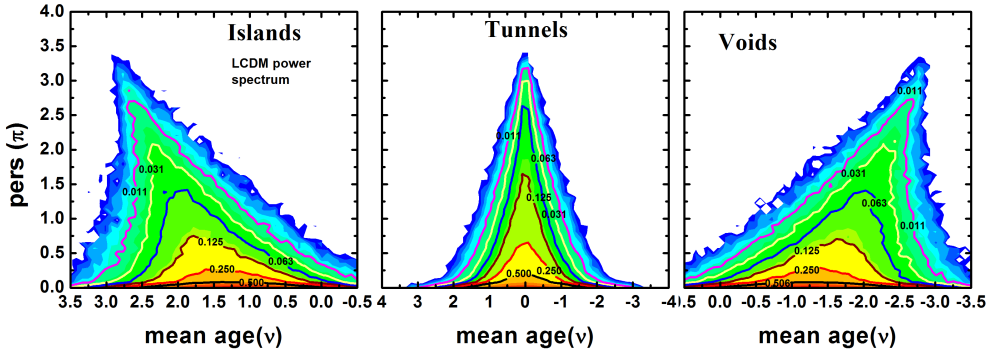


Figure 8 Intensity maps of islands, tunnels and voids (from left to right) in Gaussian random fields for the LCDM power spectrum. Intensity I is defined as the fraction of total objects (for each dimension separately) in a grid cell. Successive contours have intensity value that differ by powers of 2.

creases, the structures become visibly larger.

In Chapter 3, we analyze the persistence topology of the models. We demonstrate that ensemble average of the persistence maps of Gaussian fields are stable and well defined. As an example, Figure 8 presents the ensembe average of 100 realizations of persistence diagrams of islands, tunnels and voids for the LCDM spectrum. We also demonstrate that the features of the persistence maps are a direct reflection of the singularity structure of the field. The characteristics of the corresponding difference maps depend strongly on the choice of the model. It may imply them to be of use for discriminating between various models. This is useful when more traditional methods might not be able to accomplish this.

Useful quantitative information is extracted from the marginal and cumulative distributions of the mean-density and persistence of the topological features. We establish that the distribution functions show a characteristic dependence on the index of the power spectrum. The marginal distribution of the mean density follows a near Gaussian distribution. The marginal distribution of persistence indicates a Poisson distribution.

We also delve into an analysis of the Betti numbers of Gaussian random fields in Chapter 3. We compare the Betti numbers with the familiar Euler characteristic curves in this context. We find that the Betti numbers show a characteristic dependence on the index of the power-spectrum. This is unike the Euler characteristic curve, that has the same shape irrespective of the spectral index. A visual impression of this may be obtained from inspecting the scaled Betti number and Euler characteristic curves for Gaussian random fields presented in Figure 9.

This difference between the features of the Euler characteristic and the Betti numbers has important repercussions for the description of the topological structure of the density field. Because the shape of the Euler characteristic is the same irrespective of the power spectrum, one might naively assume the topology of all the models to be the same. On the other hand, the Betti numbers overlap across a significant range of density thresholds. The extent of overlap increases with decreasing spectral

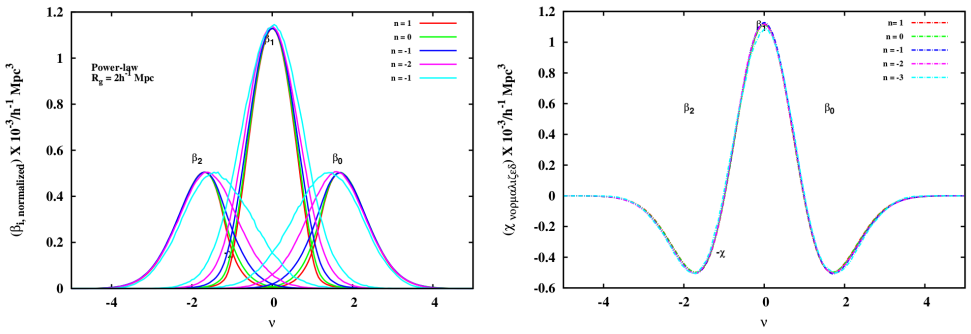


Figure 9 The curves of the scaled Betti numbers (left) and the Euler characteristic (right) for Gaussian random fields specified by a power law power spectrum. The curves are drawn for spectral indices $n = 1, 0, -1, -2 \& -3$. The curves are scaled with respect to the amplitude of the curve for the $n = 0$ model. The Betti numbers show a characteristic dependence on the choice of the power-spectrum, unlike the Euler characteristic curve. This means that the Betti numbers are sensitive to the index of the power spectrum, while the Euler characteristic is not.

index. This indicates that the topological structure of the density field is a mixture of the various kind of topologies at any given threshold. The extent of mixing increases with decreasing spectral index, there by illuminating the inherently different topological structure of the different models. This information is not revealed by the Euler characteristic curves.

5.3 Homology and persistence of cellular distributions

To model the web like patterns of the large scale matter distribution in the Universe, we resort to the *Voronoi models* (van de Weygaert 1991). Through this, we seek to identify the topological signatures of the various morphological features of the cosmic web. These models use Voronoi tessellations as a template for the distribution of matter and related galaxy population, and its subsequent evolution. The *Voronoi element models* are characterized by a dominating presence of a single morphological element. The *Voronoi evolution models* mimic the evolution of the Megaparsec universe towards a web like pattern. The models we use here are considerably sophisticated, and represent a rather realistic depiction of the cosmic web in void-dominated cosmologies. The top and the bottom panels of Figure 10 illustrate the density field rendering of a single realization of the Voronoi element and the Voronoi evolution models.

For the single component models, we demonstrate that the dominating presence of different morphologies is reflected in the features of the intensity maps in different dimensions. Models with particles exclusively distributed in and around the nodes of the underlying Voronoi skeleton exhibit the presence of numerous isolated clusters. This is captured by the high persistent isolated clouds in the 0-dimensional intensity maps. Models with particles exclusively distributed around the edges of the Voronoi skeleton mimic the filamentary structures the web. Interconnected filamentary patterns exhibit the presence of well formed loops. This is reflected in a

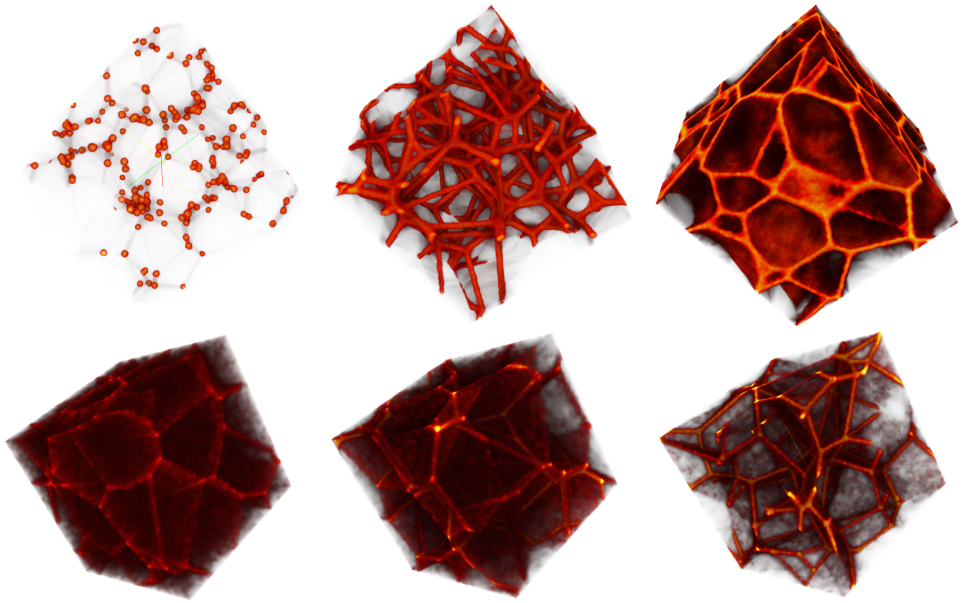


Figure 10 Density field corresponding to the Voronoi element models (top panel) and the Voronoi kinematic models (bottom panel). From left to right, the element models consist predominantly of clusters, filaments and walls. The kinematic models evolve from a stage where most of the matter is in the field, to a stage where most of the matter is confined to clusters and filaments.

high persistence cloud in the 1-dimensional intensity maps. Models with particles exclusively distributed around the faces of the Voronoi skeleton mimic the wall-like components of the web. These models exhibit the presence of fully enclosed voids. This aspect is captured by isolated high persistence clouds in the 2-dimensional intensity maps. We also demonstrate that the dominating presence of a single morphology is reflected in the presence of a single well defined mode in the graphs of Betti numbers in different dimensions.

For the Voronoi evolution models, we demonstrate that the topological characteristics of the different morphologies in the distribution – clusters, filaments and walls – are well segregated in the intensity maps. Additionally, the aspect of evolution is captured by the evolving features of the intensity maps. As an example, the left panel of Figure 11 presents the intensity maps for the islands in the highly evolved realizations of the Voronoi kinematic models. The map has four distinct features, indicating a presence of four different morphologies, namely clusters, filaments, walls and voids. The presence of multiple and evolving morphologies is also reflected as multiple well-defined modes in the Betti number curves. Figure 11 presents the Betti number curves for three different stages of the Voronoi evolution model (evolution proceeds from Stage 1 to Stage 3). The presence of multiple modes, as well as an increase in the number of modes for more evolved models is evident.

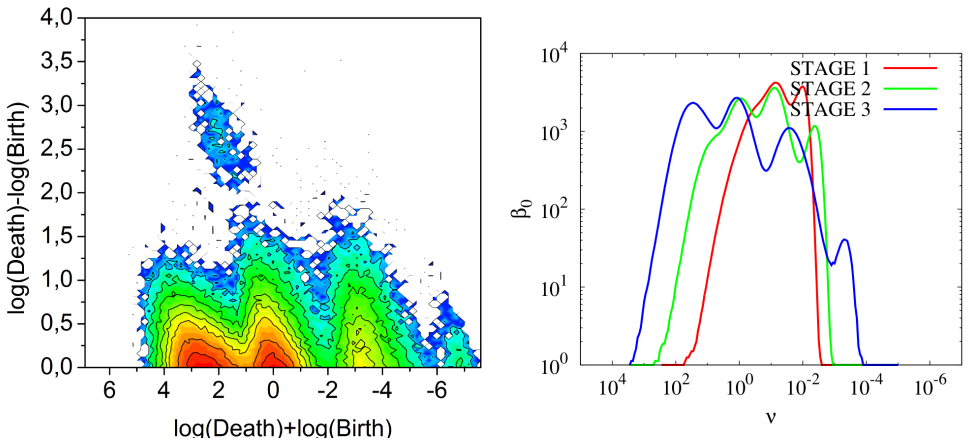


Figure 11 Left: 0-dimensional intensity maps for the highly evolved stage of Voronoi kinematic model. Right: Betti numbers as a function of density threshold for three different stages of evolution for the Voronoi kinematic models.

5.4 Homology and persistence of hierarchical distributions

To study the signature of multiscale distributions, we turn our attention to the Soneira-Peebles distribution (Soneira & Peebles 1978). It is an analytic self-similar spatial point process defined for the purpose of modeling the galaxy distribution. The model involves a hierarchy of structures of varying densities and characteristic scales. It is constructed by placing a number of children spheres at random locations inside a parent sphere, which is itself randomly placed in $\mathbb{R}^3$. The higher level spheres correspond to high density structures of small scale and the lower level spheres correspond to low density structures of larger scale. As each sphere is constructed in the same way, the resulting point distribution is self-similar and forms a bound fractal. An important property of the model is that its statistical properties can be fully and analytically evaluated. This concerns its power-law two-point

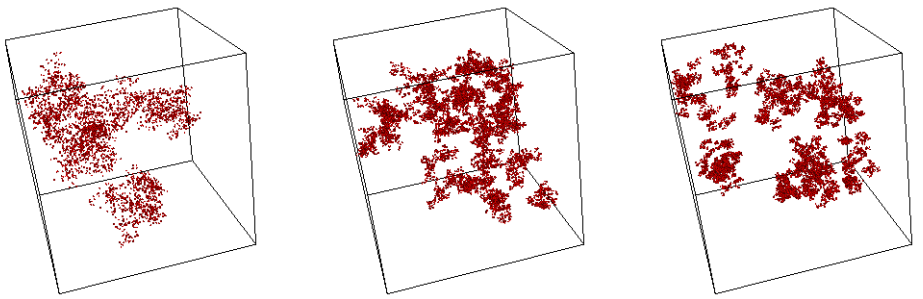


Figure 12 Particle distribution in a sample realization of the Soneira-Peebles model. Most of the volume is empty due to the high concentration of particles in a small volume.

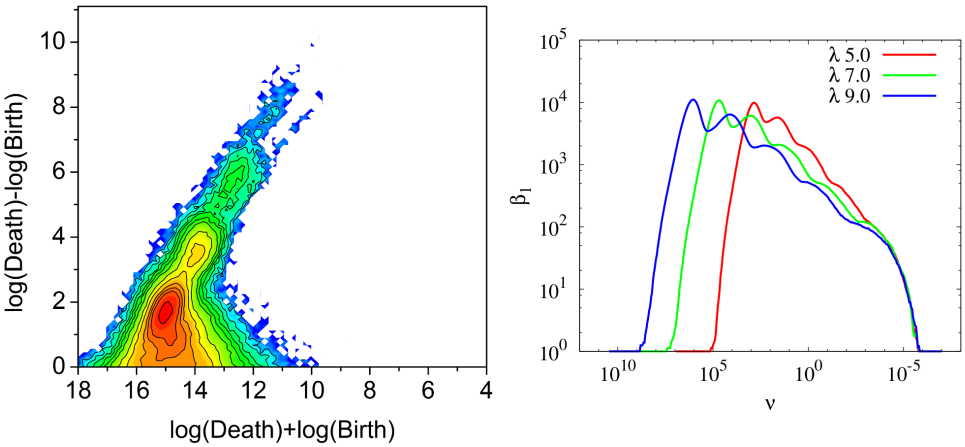


Figure 13 Left: Intensity maps for the Soneira-Peebles model with a high concentration factor. Right: Betti numbers of Soneira-Peebles model for models with different concentration factors.

correlation function, correlation dimension and its Hausdorff dimension. Figure 12 presents the particle distribution in a sample realization of the model, with three different concentrations of particles. The number of particles is the same in all the realizations.

We demonstrate that the levels of hierarchy are well manifested in the intensity maps. In the right panel of Figure 13, we present the intensity map of a typical Soneira-Peebles model. The map exhibits a presence of a number of distinct clumps. These clumps are placed at increasingly higher persistence values. The number of clumps is related to the number of level of hierarchies in the distribution.

The Betti numbers also capture the various levels of hierarchies in the distribution. We see this in the right panel of Figure 13. The multiple peaks in the graphs capture the number of levels in the hierarchy. For high concentration factors, the peaks are well segregated, denoting that the levels of hierarchy are well segregated as well. For low concentration factors, the levels are not well segregated. Hence, the particles tend to have a diffused distribution. This is reflected in the relatively low number of distinct peaks in the Betti number curves.

6 Topology based visualization of Cosmic filaments: Felix

The final part of this thesis introduces and describes a technique for the identification of filaments based on the topological characteristics of the density field. Exploring the filamentary patterns of the cosmic web is challenging because of the large range of the spatial scales and density range it exhibits. A proper characterization should also account for the hierarchical nature of structures, which adds considerable challenge to the task. The distinction between noise and significant structures is often ill-defined, and at occasions noise may be confused with genuine structures in the hierarchically evolved mass distribution. This problem is more pronounced when one studies the properties of tenuous filaments and walls in low density void-

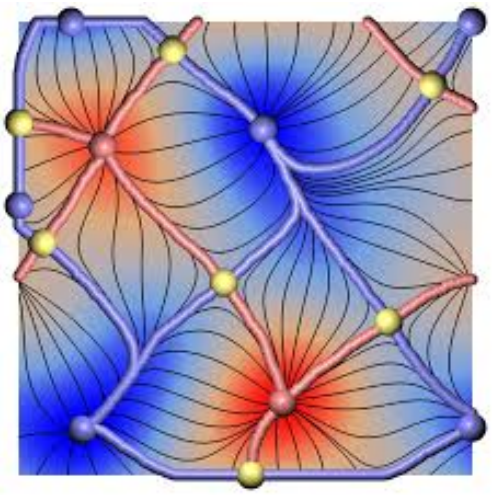


Figure 14 Morse-Smale complex of a 2D scalar function. The critical points are marked by solid spheres. Pink denotes the maxima, yellow denotes the saddle points and blue denotes the minima. The thin gray lines originating and terminating at the various critical points are the integral lines of the function. The thick gray tubular structures denote the separatrices, which form the boundary of two adjacent ascending or descending manifolds.

like regions. A paramount requirement for understanding the dependence of formation and evolution of galaxies on the properties of the local environment is the ability to classify these properties of the local environment accurately first. Such properties include the density characteristics, as well as the morphological classification.

Morse theory and related formalisms can be harnessed to devise methods of partitioning a manifold (Forman 2002; Edelsbrunner et al. 2001). This particularly depends on the geometry of the space in the neighborhood of the critical points of f. Such a partitioning of the domain is achieved by exploring the features of the gradient of the scalar function on the manifold.

An important ingredient in understanding how the partitioning of the manifold may be achieved is the concept of *integral lines*. An integral line is a path where the tangent of the path is parallel to the gradient ∇f at every point along the path. As an example, integral lines of a 2D scalar function are represented in thin gray lines in Figure 14.

The integral lines define the *ascending* and *descending* manifolds of a critical point. The set of all integral lines that originate at the critical point p together with p is called the *ascending manifold* of p. Similarly, the set of all integral lines that terminate at the critical point p together with p is called the *descending manifold* of p. The ascending manifolds (similarly, the descending manifolds) of all critical points partition the domain. The *Morse-Smale complex* is constructed by decomposing the manifold into cells formed by the collection of integral lines that share a common source and a common destination (Forman 2002; Edelsbrunner et al. 2001). This can only happen if the ascending and descending manifolds of all pairs of critical points intersect only transversally. A function f satisfying this condition is called a *Morse-Smale function*. Figure 14 presents the Morse-Smale complex of a 2D scalar function. The maxima of the function are marked in pink, the saddle points in yellow, and the minima in blue. The tubular elements represent the arcs of the Morse-Smale complex

that partition the domain.

The various elements of the cosmic web have natural geometric associations with the elements of the Morse-Smale complex constructed from cosmic density fields. Clusters can be identified as the local maxima of the field. This is because mass flows into the clusters from all the directions. This corresponds to the fact that the gradient of the field increases in the drection of the local maximum in the corresponding Morse-Smale complex. The filaments can be identified as the ascending manifold of 2-saddles. The ascending manifolds of a 2-saddle connect two neighboring maxima, and is a 1-dimensional element. The gradient along the ascending manifold of a 2-saddle increases in the direction of the maxima. This has natural associations with the cosmic filaments that are also quasi-linear structures connecting the clusters, and serve as channels through which mas flows into the clusters. Voids can be identified as the ascending manifolds of minima, from where all matter flows out. The walls can be identified as the descending manifolds of 1-saddle that separate these voids, and are 2-dimensional in nature.

We describe Felix[6], a topology based visual interactive framework to extract filamentary structures from a hierarchy of Morse-Smale complexes of the density field. The filaments in Felix are parameterized by the density values of the maxima and the 2-saddle that define them. We allow for a user-defined choice of density range, thereby providing greater flexibility and accuracy in detecting the features of interest. This is unlike the strategy of persistence based simplification as implemented in DisPerSe (Sousbie 2011), where the low density features are necessarily filtered away.

As an example, we investigate the nature of filaments in three different density regimes from the LCDM simulations. The first concerns filaments in the high density regions around compact dense clusters, which are known to function as the transport channels along which matter moves into the clusters. The second regime concerns the tenuous low-density filaments found in low-density void regions. The third regime concerns the filaments that stretch from cluster like regions all the way down to void like regions. Figure 15 presents an example of the three classes of filaments extracted from a cosmological simulation.

Using Felix, we classify galaxies as cluster/filamentary or not in test models, for which the true classification of the galaxies is known a-priori. By comparing the classification as achieved by Felix with the a-priori known classification, we demonstrate that we are able to recover the true classification with high efficiency. We compare Felix with a few other geometric and topological methods, namely DisPerSE, MMF/Nexus and SpineWeb (Aragón-Calvo et al. 2010). We demonstrate in the process that we perform equally good, if not better, in most conditions. Using the Voronoi models as a test, we demonstrate that a persistence based simplification of the Morse-Smale complex, as implemented in DisPerSE, is not a sufficient mechanism to extract the desired filaments.

[6]The name Felix is formed from an abbreviation of *Fil*ament *ex*plorer.

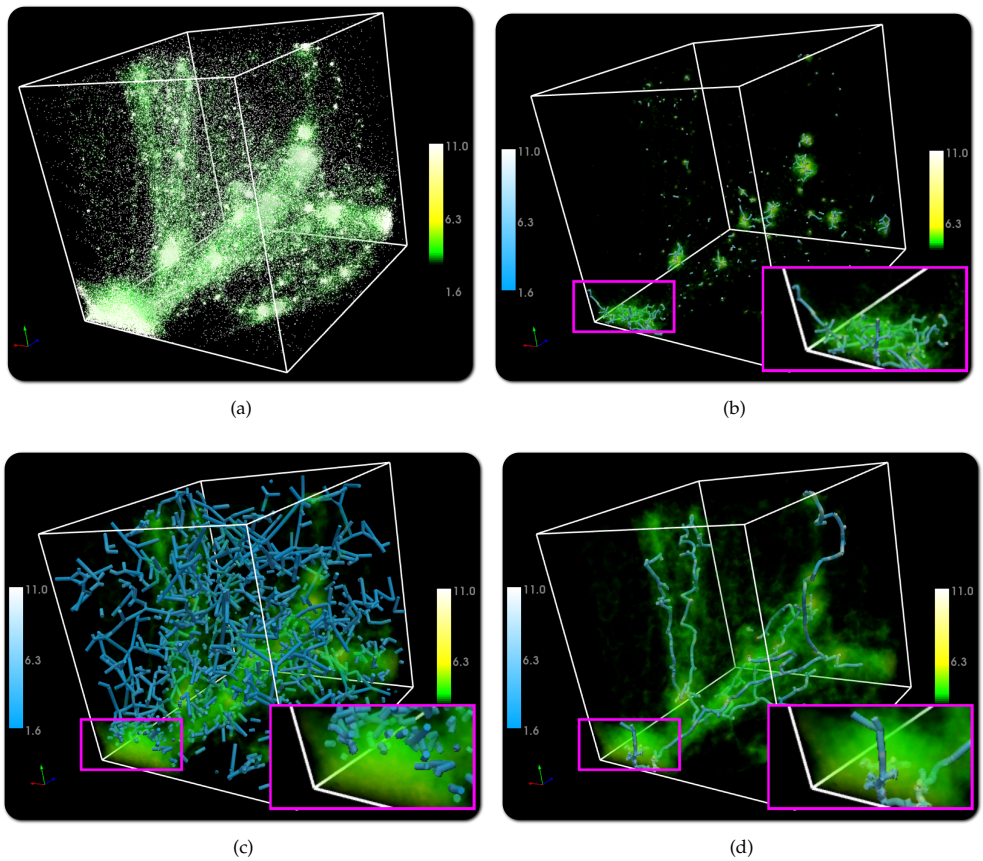


Figure 15 Three classes of filaments extracted from the Cosmogrid dataset. (a) Particle distribution shown along with a volume rendering of the DTFE density. (b) Filaments within cluster like regions. The highlighted region shows the retention of intricate topological structures within a large cluster region. (c) Filaments within void like regions extracted with parameters. Shown in the inset is a cluster like region within which filamentary structures are filtered away. (d) Filaments that stretch from cluster like regions all the way down to void like regions.

7 Conclusion

Topological data analysis, in particular persistence based analysis, of structural patterns has gained interest across various disciplines like medical imaging, cartography, agriculture etc. This is motivated by the similarity in the nature of the problems approached. Across all the disciplines, data sets are getting bigger and more complex, whether it is medical records, genomic sequencing, neural networks in the brain, astrophysics, historical archives, or social networks. Topology may be the ideal tool for analyzing patterns in these huge datasets with high dimensionality. This is because topology perhaps is the only formalism that has the ability to rigorously define shapes and patterns in any dimension.

This thesis is a culmination of an interdisciplinary collaboration between the fields of cosmology, mathematics and computer science. On the cosmological side, it

has aimed to expand the topological description of cosmic density fields using Morse theory, homology and persistence. On the mathematical side, it has attempted to explore the general properties and characteristics of persistence homology and persistence diagrams.

7.1 Future directions

At the end of this thesis, there are a number of follow up investigations that come to mind. We list here a possible, but not exhaustive, set of directions that may be promising.

Statistics of persistence. We have established empirically that the ensemble average of persistence diagrams of stochastic processes are well-defined. An analytic probabilistic and statistical description of persistence will be a natural and important extension of the framework.

Persistence (Homology) of the large scale Universe. In this thesis, we presented a topological characterization of Gaussian random fields, with a view to understand the nature of fluctuations in the primordial Universe. Subsequently, we analyze the Voronoi models that mimic aspects of the matter distribution in the large scale Universe. A logical and important extension of this would be to investigate the topology of the genuine distributions of dark matter and halos in LCDM simulations.

Persistence characterization of the anisotropies in CMB. A persistence based hierarchical characterization of the anisotropies in the Cosmic Microwave Background, as provided by the latest measurements from the PLANCK satellite provides an interesting challenge Persistence has the potential to shed light into the hierarchical nature of primordial fluctuations. It may also be interesting to investigate the CMB with a view to isolate signatures of primordial non-Gaussianities. Using existing methodologies, the PLANCK team reports the absence of statistically significant non-Gaussian signals in the CMB maps. The hierarchical description of persistence has the potential to probe deeper in this direction.

Filament catalogues of the large scale Universe. Having tested the robustness of the filament finding software on test models and simulations, an important aspect to follow up is to produce filamentary catalogues from simulations as well as observation. These catalogs will have an option of querying for the properties of the associated galaxies they host, in view of investigating the formation and evolution of galaxies vis-a-vis the large scale environment they reside in.

Hierarchical characterization of Cosmic Voids. 2-dimensional persistence diagrams, that capture the formation and evolution of topological voids, can potentially provide a powerful tool for the characterization of the hierarchical aspects of cosmic voids. This is especially interesting, in view of the fact that topology, perhaps, is the only method that defines a void uniquely and unambiguously, without a choice of any free parameter. It has theoretically been suspected for long, and confirmed recently observationally that the cosmic voids form and evolve hierarchically. Stud-

ies have revealed cosmic voids are highly sensitive to dark energy and may contain information about modified gravity.

Characterization of the epoch of reionization. Due to its hierarchical nature, persistence is naturally tailored towards the study of reionization and the evolving network of ionization bubbles in the Epoch of reionization, marking the onset of formation of stars and galaxies. 0-dimensional persistence diagrams are optimally suited for capturing the evolving network of ionization bubbles, which are essentially isolated objects undergoing mergers and possibly more complex procedures. Additionally, studying the characteristics of the 1-dimensional diagrams will reveal details about the percolation properties of the distribution at that epoch. The 2-dimensional diagrams have a potential to reveal the evolving hierarchical network of voids in their infancy.

Samenvatting

1 Het kosmisch web

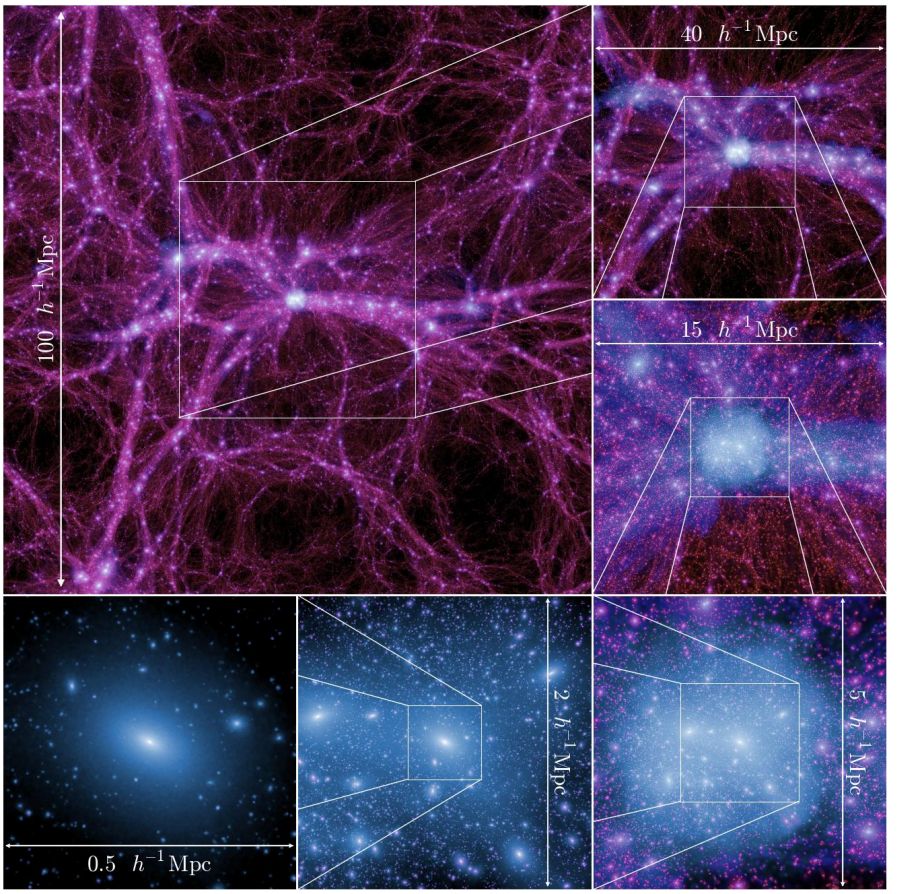
Hoewel materie in de kosmos op schalen van gigaparsecs[7] gelijkmatig verdeeld lijkt, is op kleinere schaal, van enkele to honderden megaparsecs, een ingewikkeld webachtig patroon zichtbaar dat bekend staat als het *Kosmisch Web*. Dat dit kosmisch web daadwerkelijk bestaat, is bevestigd door de analyse van een scala van onderzoeken naar de roodverschuivingen van melkwegstelsels, zoals de CfA en CfA2 surveys (de Lapparent et al. 1986; Huchra & Geller 1982), de 2dFGRS survey, (Colless et al. 2003), de SDSS survey (Abazajian et al. 2003) en de 2MRS survey (Huchra et al. 2005; Skrutskie et al. 2006).

Zware computersimulaties van het heelal op schalen van megaparsecs tot gigaparsecs laten ook het bestaan van dit kosmisch web zien. (Springel 2008). Figuur 1 geeft een impressie van het Kosmisch Web op verschillende schalen, berekend door de allernieuwste Millenniumsimulatie.

Het kosmisch web bestaat uit grote, in viriaal evenwicht verkerende *clusters* die duizenden sterrenstelsels bevatten, alsmede grote hoeveelheden donkere materie en zeer heet gas. Uit deze gevirialiseerde clusters verspreiden zich gigantische *filamenten*. Deze filamenten, die ook weer duizenden melkwegstelsels, donkere materie en gas bevatten, fungeren als transportkanalen voor massastromen in het Heelal. Daarnaast bestaat het kosmisch web uit grote, platte objecten, aangeduid als *walls (wanden)*. Deze clusters, filamenten en wanden zijn vervlochten rondom uitgestrekte lege ruimte: de *kosmische voids*.

Volgens het gangbare paradigma voor de vorming van structuur in het heelal, het *zwaartekrachtsinstabiliteitsscenario*, groeit de kosmische structuur uit microscopische oneffenheden in de dichtheid in het zeer jonge heelal. De temperatuurfluctuaties

[7] De parsec is de gebruikelijke maat voor afstanden in het Heelal. Eén parsec is gelijk aan 3.26 *lichtjaar*, de afstand die licht aflegt in één jaar. Een megaparsec is gelijk aan een miljoen parsec, een typische maat voor de grootte van de groteschaalstructuren in het Heelal.



Figuur 1 Opeenvolgende vergrotingen van een deel van de Millennium-II simulatie laten de hiërarchische structuur van het heelal zien. Dit is een donkere materiesimulatie die het anisotropische en meerschalige karakter van de evoluerende massaverdeling in het heelal blootlegt. Het aantal gebruikte deeltjes is meer dan 10 miljard. Het grootste plaatje (linksboven) toont een $15h^{-1}$Mpc dikke plak uit de $100h^{-1}$Mpc grote simulatiekubus op roodverschuiving nul, gecentreerd op de zwaarste halo in de simulatie. De clusters, filamenten, wanden en grote leegtes zijn duidelijk zichtbaar. De volgende kaders laten vergrotingen zien van kleinere gebieden. Op alle schalen ziet men een rijke verscheidenheid van substructuur. Met dank aan Boylan-Kolchin et al. (2009) voor de figuur.

in de kosmische achtergrondstraling (Planck Collaboration et al. 2013b) suggereren dat deze oneffenheden beschreven kunnen worden door een homogeen en isotroop, ruimtelijk Gaussisch proces. Theorieën over het vroege heelal voorspellen dit soort Gaussische oneffenheden als een natuurlijk nevenprodukt van een eerdere inflationaire fase van het heelal.

Het megaparsec-grote kosmische web is een van de frappantste voorbeelden van complexe geometrische patronen in de natuur en welzeker het grootste. De eigenschappen van het kosmisch web zijn buitengewoon anisotroop. Zij kenmerken zich

door een grote verscheidenheid aan vormen en dimensies. Clusters zijn gevirialiseerde, bijna-bolvormige objecten, die de verst geëvolueerde massaverdeling vertegenwoordigen. Filamenten en wanden zijn asymmetrisch en in het algemeen in één of twee dimensies groter dan in de andere dimensies. Het meest prominente morfologische bestanddeel van het web, met betrekking tot het volume, zijn de bijna lege, enorme kosmische voids.

De morfologische bestanddelen van het web tonen ook een andere asymmetrie, met betrekking tot de massafractie en in beslag genomen volume. De clusters hebben de hoogste massafractie, maar nemen het kleinste volume in beslag. Daarentegen bevatten de voids relatief de minste massa, terwijl zij het grootste volume in beslag nemen. Dit is tegenstrijdig met de initiële Gaussische randvoorwaarden, die symmetrisch zijn in hoeveel relatief volume overdichte en onderdichte gebieden in beslag nemen.

Alle bestanddelen van het web zijn alomtegenwoordig op alle schalen en dichtheden. Dit is de meest duidelijke manifestatie van het hiërarchische karakter van de evolutie van de massaverdeling in het heelal. Ten slotte laat het web ook kenmerkende complexe verbindingskarakteristieken zien. De laag-dimensionale morfologische bestanddelen van het web vormen de grenzen van de hoger-dimensionale bestanddelen, resulterend in een doordringend netwerk van complexe en onderling verbonden patronen.

Het is een grote uitdaging om de structuur, geometrie en topologie van het kosmisch web te beschrijven. De overweldigende complexiteit van individuele structuren alsmede hun onderlinge verbindingen en afwezigheid van interne symmetrie, het intrinsieke meerschalige karakter van het web en het grote spectrum van dichtheden in de massaverdeling werken het gebruik van simpele voor de handliggende methodes tegen.

Hieronder geven we een korte samenvatting van de analysemethodes van het kosmisch web.

2 Kenschetsing van het kosmisch web

Het streven om de massaverdeling in het heelal te beschrijven en te karakteriseren vormt al decennialang een uitdaging. Een kwantitatieve karakterisering van de complexe en rijke ruimtelijke patronen in het heelal is zeer belangrijk voor het vergelijken en valideren van verschillende theoretische modellen met de werkelijke massaverdeling in het heelal. De 2-puntscorrelatiefunctie, of autocorrelatiefunctie, vormt de voornaamste steunpilaar voor statistische analyse in de kosmologie (Peebles 1980). Voor een uitstekend en uitgebreid boek over de statistische karakterisering van de grootschalige massaverdeling raadplege men Martínez & Saar (2002). Echter, de autocorrelatiefunctie is niet gevoelig voor de fasen van de Fouriercomponenten. Hierdoor is het mogelijk dat twee massaverdelingen met verschillende geometrische en morfologische patronen dezelfde correlatiefunctie hebben. Men kan hogere-orde correlatiefuncties gebruiken om de statistiek van de massaverdeling te beschrijven. Echter, deze correlatiefuncties hebben maar beperkt nut, omdat de onzekerheden in de meting onevenredig groot worden met oplopende orde.

In het afgelopen decennium is men tot het inzicht gekomen dat men voor een vol-

ledige analyse van de rijke structuren in het heelal rekening moet houden met zowel de geometrische als de morfologische aspecten van de elementen van het kosmisch web.

De statistische maatregelen hiervoor omvatten de analyse van de globale verbondenheid van het kosmisch web, door middel van *percolatieanalyse* (Shandarin & Zeldovich 1983) en de *minimaal opspannende boom* (Barrow et al. 1985) van de melkwegstelselverdeling. Hoewel percolatieanalyse en opspannende bomen relatief gevoelig zijn voor de onderliggende structuur, zijn het strikt globale eigenschappen die nalaten de lokale variatie in vorm te beschrijven.

Noemenswaardig zijn ook de modellen en methodologieën die uitgaan van statistische en geometrische overwegingen voor de afbakening van kosmische filamenten. De meest prominente hiervan zijn het *snoepmodel*, dat een gemarkeerd puntproces gebruikt om de filamenten af te bakenen (Stoica et al. 2005; Tempel et al. 2013) en het terugvinden van filamenten door ze te herkennen als de mediale as van een bepaald puntproces (Genovese et al. 2010).

Ook morfologische methodes, gericht op het vaststellen van de lokale vorm door te kijken naar de eigenschappen van de eigenwaarden van de Hessiaan van verschillende verdelingen die voorkomen in de kosmologie, hebben in de voorhoede gestaan bij het afbakenen van het kosmisch web. Deze methodes omvatten onder andere de eigenwaarden van de Hessiaan van een dichtheidsverdeling (Aragón-Calvo et al. 2007b; Cautun et al. 2013), die van de getijdentensor of de snelheids/shear-tensor (Tempel et al. 2014).

Een variatie op deze methodes evalueert dit verder op meerdere schalen, door het uitvoeren van dezelfde procedures op een stapel afbeeldingen met verschillende schalen, die gemaakt worden door de oorspronkelijke afbeelding te versmeren met een voor de schaal geschikt filter (Aragón-Calvo et al. 2010; Cautun et al. 2014).

Vormbepaling op basis van morfologie is nauwgerelateerd aan vormbepalingsmethodes gebaseerd op topologische afwegingen. Beide methodes identificeren een morfologisch element middels de eigenschappen van de Hessiaan van de massaverdeling, ofwel voor uitgestrekte objecten (morfologische methoden) ofwel voor gelocalizeerde punten in de ruimte (topologische methoden).

Behalve dat zij bruikbaar zijn voor het afbakenen van de structuren in het web, zijn topologische methodes ook belangrijk vanuit het oogpunt van onderzoek naar de inherente structuur en verbindingen tussen de verschillende structurele elementen die de complexe patronen van het kosmisch web vormen. In de volgende paragraaf presenteren we een kort verslag over topologie, in het bijzonder voor kosmologische toepassingen. Dit is het hoofdthema van dit proefschrift.

3 Topologie van de massaverdeling in het Heelal

De topologie is dat deel van de wiskunde dat zich bezighoudt met eigenschappen van ruimtes die behouden blijven onder continue vervormingen zoals uitrekken (comprimeren) en ombuigen, maar niet scheuren of aan elkaar lijmen. Het bevat eveneens onveranderlijke eigenschappen zoals samenhang en rand.

Dit proefschrift houdt zich bezig met de topologische analyse van scalaire velden die verschillende aspecten van de kosmische massaverdeling beschrijven. Onze stra-

tegie is om zulke functies of velden op een variëteit te definiëren en de geïnduceerde topologie van deze functies op de variëteit te bestuderen. De variëteit in dit proefschrift is ofwel $\mathbb{R}^3$ danwel $\mathbb{T}^3$. Deze laatste wordt gevormd door de tegenover elkaar liggende vlakken van een kubus aan elkaar te lijmen.

Wij bestuderen de topologie van de variëteit door middel van het traceren van topologische veranderingen voor een groeiende superlevelverzameling [8]. De kritieke punten van het scalaire veld zijn verantwoordelijk voor de topologische veranderingen op de variëteit [9] (Milnor 1963). Het karakter en de verdeling van kritieke punten levert inzichtelijke informatie over de eigenschappen van het veld zelf op. Bijvoorbeeld, een symmetrische verdeling van kritieke punten rond het gemiddelde van het dichtheidsveld suggereert de aanwezigheid van een Gaussisch veld.

In het kader van het kosmisch web is het bekend dat individuele structuur elementen met verschillende dimensies met elkaar verbonden zijn op een complexe en hiërarchische wijze. De topologie biedt de mogelijkheid om deze vormen met verschillende dimensies te karakteriseren. We kunnen ook de complexe structurele elementen beschrijven, en hoe ze verbonden zijn in het web, met behulp van topologische concepten. De analyse van de topologische samenhang helpt ons om beter inzicht te krijgen in hoe de ontwikkeling van de grote schaal-structuur van het kosmische web ontstaat wanneer naburige structuren van kleinere schaal met elkaar interageren. Samensmelting en mogelijk fragmentaties zijn de meestvoorkomende voorbeelden van dergelijke interacties.

Omdat de topologie ongevoelig is voor continue vervormingen zoals uitrekken en buigen, verwacht men dat topologische grootheden ook betrekkelijk ongevoelig zijn voor systematische effecten zoals niet-lineaire zwaartekrachtsevolutie, melkwegstelsel biasing en roodverschuivingsruimtevervorming (Park & Kim 2010). Bijgevolg verwacht men dat de topologie van het oorspronkelijke veld en het grootschalige lineaire veld op latere tijd van dezelfde aard zijn (Gott et al. 1986).

Topologische studies worden sinds tientallen jaren gebruikt om de kosmische massaverdeling the beschrijven. De eerste onderzoeken richtten zich met name op de beschrijving van de massaverdeling door het geslacht en de Eulerkarakteristiek te bepalen van iso-oppervlakken van de massadichtheid als functie van drempeldichtheid (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Gott et al. 1986; Hamilton et al. 1986).

Later kwam meer onderscheidende informatie beschikbaar met de introductie van Minkowskifunctionalen (Mecke et al. 1994). Het is belangrijk op te merken dat deze onderzoeken grotendeels een heuristisch karakter hadden, en zich met name richtten op het onderscheid maken tussen modellen. In wat volgt gaan we in op de belangrijkste topologische methodes: geslacht, Euler- en Minkowskifunctionalen.

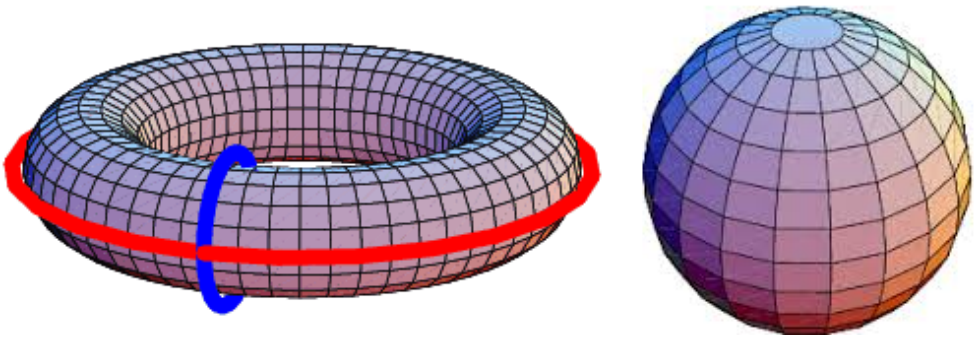
3.1 Geslacht

Voor een samenhangend, oriënteerbaar oppervlak is het *Geslacht*[10] gedefinieerd

[8] Voor een gegeven functiewaarde ν is de *superlevelverzameling* of het *excursiegebied* dat deel van de variëteit dat boven een bepaalde drempelwaarde ν ligt. De rand van dit gebied heet de *levelverzameling* of *excursieverzameling*.

[9] Een kritiek punt van een functie is de plaats waar de gradiënt van de functie verdwijnt, dus $\nabla f = 0$.

[10] Voor consistentie is het belangrijk om op te merken dat de definitie van het geslacht g in kosmologi-



Figuur 2 Illustratie van het concept geslacht. Het geslacht is het aantal onafhankelijke gesloten krommes dat getekend kan worden op een oppervlak, zonder dat delen van dit oppervlak niet langer samenhangend zijn. Voor een torus kunnen twee onafhankelijke gesloten krommes getekend worden zonder dat het oppervlak niet langer samenhangend is (getekend in rood en blauw). Het geslacht is daarom 2. Het is niet mogelijk om een gesloten kromme op het oppervlak van een bol te tekenen zonder het oppervlak verbroken wordt. Dus, het geslacht van een bol is 0.

als het maximale aantal onafhankelijke, simpele, gesloten krommes op het oppervlak, zonder dat de variëteit onsamenhangend wordt.

We geven een intuïtieve interpretatie van geslacht door de illustratie van een torus en een bol in Figuur 2. Er kunnen twee onafhankelijke gesloten krommes getekend worden over het oppervlak van een torus zonder dat de variëteit daardoor losgekoppeld wordt. Deze krommes zijn getekend in rood en blauw. Dit vertelt ons dat het geslacht van een torus 2 is. Anderzijds is het evenwel onmogelijk om een gesloten kromme over het oppervlak van een bol te tekenen, zonder dat de bol onsamenhangend wordt. Vandaar dat het geslacht van een bol 0 is.

3.2 Eulerkarakteristiek

Het geslacht is nauw gerelateerd aan een grootheid die misschien wel het meest aan de grondslag ligt van de topologie van oppervlakken – de *Eulerkarakteristiek*. Laat een gesloten orienteerbaar oppervlak $\mathbb{M}$ een geslacht g hebben. Dan zegt de *Eulerformule* dat

$$\chi(\mathbb{M}) = V - E + F, \tag{5.11}$$

waarbij

$$\chi(\mathbb{M}) = 2 - 2g \tag{5.12}$$

de Eulerkarakteristiek is, soms ook wel bekend als de *Euler-Poincarékarakteristiek*. V, E en F zijn het aantal vertices, ribben en vlakken gebruikt om het oppervlak te tri-

sche studies verschilt van de wiskundige definitie van geslacht G, met als verschil het aantal aaneengesloten gebieden:$g = G - c$.

anguleren. In meer algemene taal, de Eulerkarakteristiek van een variëteit is het totaal aantal even-dimensionale simplices minus het totaal aantal oneven dimensionale simplices dat nodig is voor de triangulatie. De Eulerkarakteristiek is afkomstig uit de analyse van veelvlakken.

Laat ons als voorbeeld de bol bekijken. Topologisch is de bol equivalent aan een tetraëder. Een tetraëder heeft 4 vertices, 6 ribben en 4 vlakken. De Eulerkarakteristiek van de bol (of van de tetrahedron) is dus 2. In feite heeft ieder convex veelvlak, zoals de kubus, octaëder, dodecaëder, dezelfde Eulerkarakteristiek als een tetraëder of een bol, omdat zij topologisch hetzelfde object zijn.

3.3 Minkowskifunctionalen

Minkowskifunctionalen zijn gerelateerd aan de geometrische eigenschappen van een variëteit. Stel dat we een vast lichaam nemen, $\mathbb{M}$, waarvan de rand een glad oppervlak ingebed in $\mathbb{R}^3$ is. Dit oppervlak kan een bol zijn, mag gaten hebben zoals de torus, en mag bestaan uit een of meerdere samenhangscomponenten met elk hun eigen gaten. We verlangen niet dat $\mathbb{M}$ samenhangend is. We definiëren $\mathbb{M}^r$ als de verzameling punten met afstand tot $\mathbb{M}$ kleiner of gelijk aan r. Voor r klein zal de rand van $\mathbb{M}^r$ glad ingebed zijn in $\mathbb{R}^3$, maar naarmate r groter wordt, zal deze inbedding singulariteiten en zelfdoorsnijdingen laten zien. Voordat dit gebeurt kan het volume van $\mathbb{M}^r$ geschreven worden als een polynoom van de derde graad in r,

$$\text{vol}\,\mathbb{M}^r \;=\; Q_0 + Q_1 r + Q_2 r^2 + Q_3 r^3\,. \tag{5.13}$$

De Q_i staan bekend als de *Minkowskifunctionalen* van $\mathbb{M}$ en zijn belangrijke begrippen binnen de integraalmeetkunde. Voor een d-dimensionale variëteit $\mathbb{M}$ zijn er $(d+1)$ Minkowskifunctionalen, $Q_k((k = 0, \ldots, d)$ (Munkres 1984). In dit proefschrift zijn we met name geïnteresseerd in $d \leq 3$. In 3D is Q_0 gelijk aan de inhoud van $\mathbb{M}$, Q_1 het oppervlak van de rand, Q_2 de totale gemiddelde kromming en Q_3 gelijk aan een derde van de totale Gaussische kromming van de rand. Volgens de stelling van Gauss-Bonnet is Q_3 gelijk aan de Eulerkarakteristiek van de rand van $\mathbb{M}$ (modulo een multiplicatieve constante).

3.4 Voorbij de Eulerkaracteristiek en Minkowskifunctionalen

Zowel geslacht en Eulerkarakteristiek, als de rest van de Minkowskifunctionalen, zijn verantwoordelijk geweest voor cruciale inzichten in de topologie en morfologie van kosmische dichtheidsvelden, zowel in modellen als uit waarnemingen. De Eulerkarakteristiek van kosmische dichtheidsvelden in het lineaire regime suggereert bijvoorbeeld de aanwezigheid van bijna-Gaussische randvoorwaarden in het vroege heelal. Toch zijn er een aantal redenen om een groter en gedetailleerder arsenaal van topologische concepten en beschrijvingen te verkennen. Allereerst merken we op dat structuren in het Heelal zich vormen en evolueren op een hiërarchische manier (Peebles 1980). Kleinere hoge dichtheidsstructuren smelten hiërarchisch samen en produceren zo grotere structuren van lagere dichtheid. Geslacht, Eulerkarakteristiek en Minkowskifunctionalen zijn niet toegerust om dit aspect te bestuderen. Een formalisme dat in staat is om de topologie in een intrinsiek hiërarchische context te bestuderen, is een handig gereedschap binnen het arsenaal van middelen om de

hiërarchische opbouw van structuur in het heelal te beschrijven.

Ten tweede vertelt de Euler-Poincaré formule ons dat de Eulerkarakteristiek een alternerende som van *Bettigetallen* is (Section 5.1.1). Dit duidt erop dat Bettigetallen in wezen rijker zijn in informatie dan de Eulerkarakteristiek en het geslacht.

Hoewel het geslacht kan onderscheiden tussen samenhangende gesloten oppervlakken in $\mathbb{R}^3$, geldt ten derde dat het dat niet kan op een 3-variëteit. In andere woorden, het geslacht is slechts gedefinieerd voor 2-dimensionale samenhangende gesloten oppervlakken en kan niet gegeneraliseerd worden naar hogere dimensies. Dit is niet wenselijk in de context van kosmische dichtheidsvelden, waar uitgebreid gebruik wordt gemaakt met 3D variëteiten.

4 Homologie, Morse-theorie and Persistentie

De uitbreiding van ons topologisch arsenaal betreft de formalismes van *homologie*, *Morse theorie* en *persistentie* in een kosmologisch verband. In dit proefschrift laten we zien dat zij het potentieel hebben om rijkere en gedetailleerder topologische informatie over de kosmische massaverdeling bloot te leggen.

4.1 Homologie en Bettigetallen

Homologietheorie is een wiskundig formalisme dat berekenbare algebrasche objecten (groepen) toekent aan topologische ruimten [11] door het vaststellen van de randen van een variëteit (Munkres 1984). Homologiegroepen geven ons een wiskundige taal voor de gaten in een topologische ruimte (Edelsbrunner & Harer 2010). Een d-variëteit kan bestaan uit topologische gaten van 0 tot $(d-1)$ dimensies. Gaten in $d < 3$ zijn intuïtief te begrijpen. Een 0-dimensionaal gat is een *kloof* tussen twee geïsoleerde onafhankelijke objecten. Een 1-dimensionaal gat is een *tunnel* waardoor men heen kan zonder de rand te raken. Een 2-dimensionaal gat is een *holte* of *leemte* (*void*), volledig ingesloten door een 2-dimensionaal oppervlak. Deze intuitieve interpretatie in termen van 'kloven' en 'tunnels' is alleen geldig voor oppervlakken ingebed in $\mathbb{R}^3$, $\mathbb{S}^3$ of $\mathbb{T}^3$.

Homologie kan ook gedefinieerd worden in termen van *cykels* die de rand van een gat vormen. In 3 dimensies is een 0-cykel een samenhangend object (en dus, de rand van een 0-gat, de kloof tussen twee onafhankelijke objecten). Een 1-cykel is een lus/cylinder die een tunnel omvat. Een 2-cykel is een *schil* rond een *holte*.

De verzameling van alle p-dimensionale cykels van een variëteit noemt men de p-de homologiegroep $\mathbb{H}_p$. De rang van deze groep is het maximale aantal *onafhankelijke* cykels, aangegeven met hun *Bettigetallen* β_p, waarbij $p = 0, \ldots, d$ (Betti 1871; Edelsbrunner & Harer 2010). In 3D is β_0 gelijk aan het aantal samenhangscomponenten, β_1 het aantal lussen rond onafhankelijke tunnels en β_2 het aantal schillen rond onafhankelijke holtes.

De Bettigetallen bevatten meer informatie over de topologie dan het geslacht en de Eulerkarakteristiek. De Euler-Poincaréformule zegt dat de Eulerkarakteristiek geschreven kan worden als een alternerende som van Bettigetallen (Edelsbrunner &

[11] Er bestaat het begrip k-samenhangendheid, $k = 0, \ldots, d$, met d de dimensie van de variëteit. Hierbinnen geldt dat 0-samenhangendheid overeenkomt met het 'gebruikelijke' begrip samenhangend.

Harer 2010; Adler & Taylor 2010):

$$\chi = \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{5.14}$$

met d de intrinsieke dimensie van de variëteit.

4.2 Morse-theorie

Morse-theorie relateert de eigenschappen van een scalaire functie f aan de topologische veranderingen die zij induceert op een variëteit waarop zij is gedefinieerd (Milnor 1963). De kritieke punten van een functie zijn verantwoordelijk voor de topologische veranderingen. Morse-theorie houd sterk verband met de verdeling van kritieke punten en de daaruit voortvloeiende singulariteitsstructuur van de functie. De *kritieke punten* zijn die locaties waar de gradiënt van f verdwijnt, ofwel, $\nabla f = 0$. De functie f heet een Morse-functie wanneer geen van haar kritieke punten ontaard is, anders gezegd, de *Hessiaan* van f, gedefinieerd als de matrix van tweedeorde partiële afgeleiden, is niet skingulier.

Met de juiste coördinaattransformatie kan een functie nabij een kritiek punt geschreven worden als een kwadratische functie zonder 'kruistermen'. De *index* van een kritiek punt is gelijk aan het aantal negatieve termen in deze uitdrukking. De index van het kritieke punt definieert de geometrie van de variëteit in de zijn nabijheid. In 2D correspondeert index 0 met een minimum, index 1 met een zadelpunt en index 2 met een maximum. In 3D correspondeert index 0 met minima, index 1 met 1-zadelpunten, index 2 met 2-zadelpunten en index 3 met een maxima.

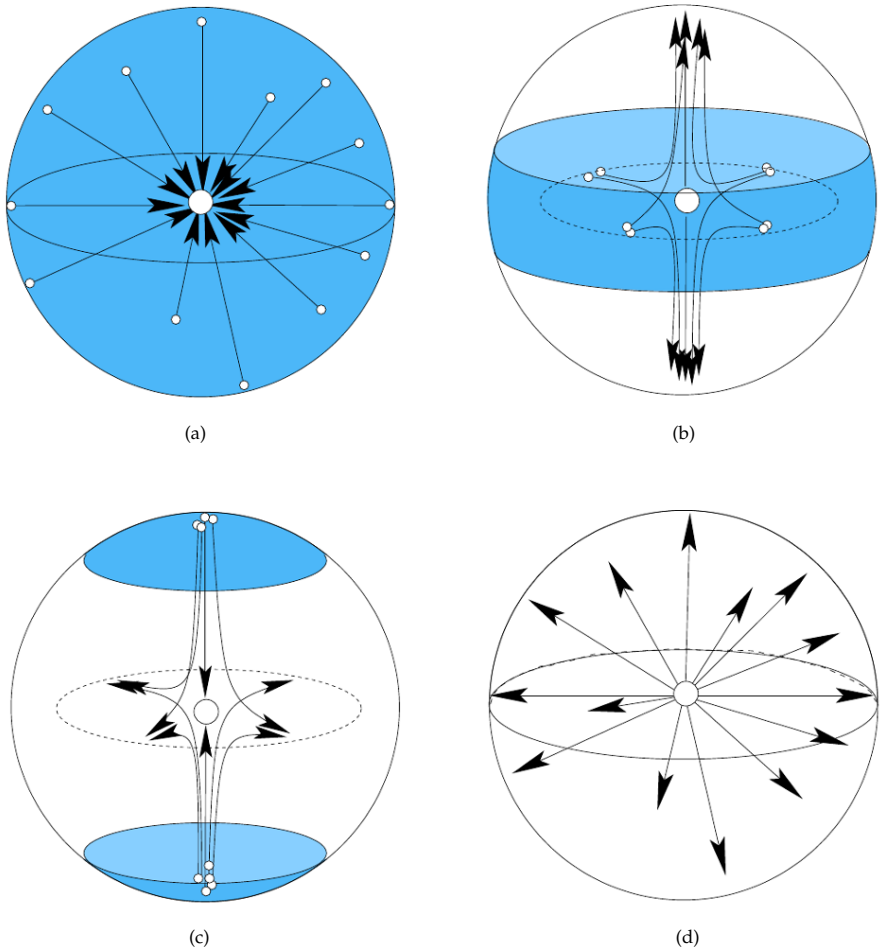
Figuur 3 laat de geometrie en het gedrag van de gradiënten in de nabijheid van verschillende soorten kritieke punten in 3D zien.

Een belangrijke stelling binnen de Morse-theorie zegt dat de kritieke punten de globale vorm en topologie van een variëteit definiëren (Milnor 1963; Edelsbrunner & Harer 2010). Veranderingen in de topologie van de superlevelverzameling van $\mathbb{M}$ komen alleen dan voor wanneer de corresponderende niveauverzameling door een kritiek punt van f heen gaat. Dit impliceert ook dat topologische veranderingen in discrete stappen optreden.

4.3 Persistentie

Afstammend van homologie en Morse-theorie is *persistentie* van nature hiërarchisch. De hiërarchische uitdrukking van topologie in persistentie wordt bewerkstelligd door de constructie van een *filtratie* van de kosmische massaverdeling. Dit kan bijvoorbeeld gedaan worden door voortdurend de superlevelverzameling te verminderen. Voor een gegeven variëteit $\mathbb{M}$ en drempelwaarde voor de dichtheid ν is de superlevelverzameling $\mathbb{M}^s(\nu)$ de deelverzameling van $\mathbb{M}$ met dichtheid $\nu_s \geq \nu$. Het verminderen van de superlevelverzameling brengt met zich mee dat een variëteit corresponderend met een hogere dichtheidsdrempelwaarde noodzakelijkerwijs ingebed is in een variëteit corresponderend met een lagere drempelwaarde.

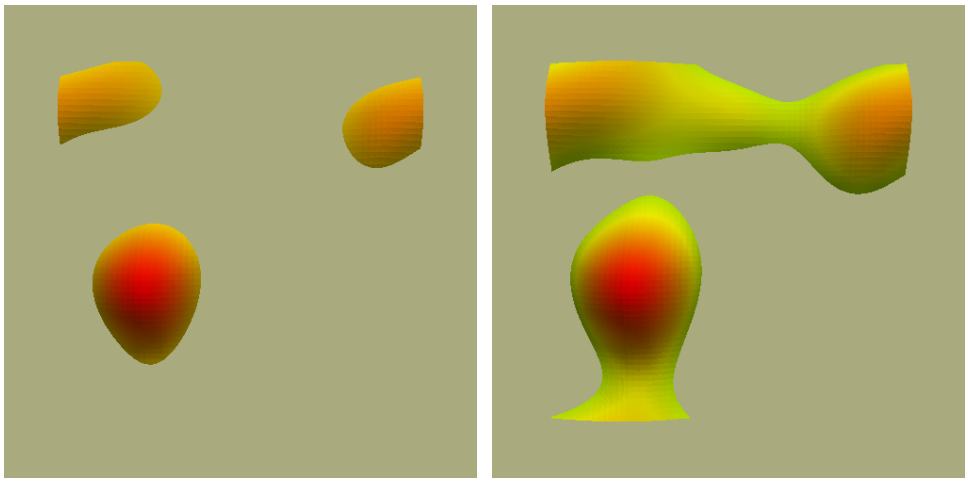
Als voorbeeld laten we de variëteiten corresponderend met twee verschillende dichtheidsdrempelwaardes van een 2D scalaire functie in Figuur 4 zien. De variëteit met de hogere drempelwaarde in paneel (a) bevat 3 afzonderlijke eilanden die rond de maxima van functie groeien. In de variëteit met de lagere dichtheidsdrempel in



Figuur 3 Kritieke punten in 3D. Paneel (a) t/m (d) laten het maximum, 2-zadelpunt, 1-zadelpunt en minimum zien. Voor een maximum is de stroom naar het maximum gericht. Voor een minimum richt de stroom zich van het minimum. Voor een 1-zadelpunt richt de stroom zich in een vlak naar het zadelpunt en van het zadelpunt weg langs een lijn. Deze richtingen zijn omgekeerd voor een 2-zadelpunt.

paneel (b) smelten twee van de drie componenten samen. Deze samensmelting komt door de introductie van een zadelpunt in de variëteit wanneer de dichtheidsdrempel verlaagd wordt. De variëteit die correspondeert met de hoogste drempelwaarde is ingebed in de variëteit corresponderend met de laagste drempelwaarde. Dit is evident wanneer we ons realizeren dat de eerste teruggevonden kan worden uit de laatste door het zadelpunt dat verantwoordelijk was voor de samensmelting te verwijderen.

De toevoeging van een kritiek punt met index p leidt ofwel tot de geboorte van

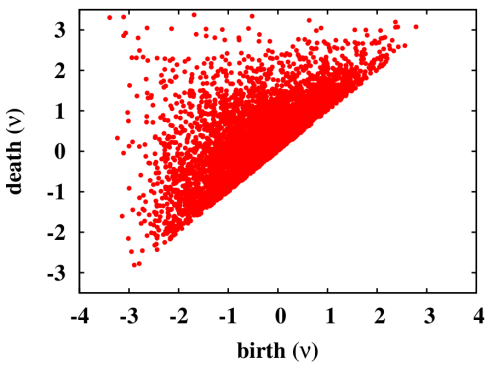


(a) (b)

Figuur 4 Dood en geboorte van eilanden en tunnels. Paneel (a) laat drie onafhankelijke eilanden zien. In paneel (b) smelten twee van deze eilanden samen, overeenkomend met de dood van een 0-dimensionaal gat. Tijdens het filteren is de variëteit in paneel (a) ingebed in de variëteit in paneel (b).

een p-dimensionaal gat of de dood van een $(p-1)$-dimensionaal gat. Elk gat kan geassocieerd worden met twee unieke functiewaardes: $f(c_b)$ geassocieerd met het kritieke punt c_b waaruit het geboren wordt, en $f(c_d)$ geassocieerd met het kritieke punt c_d dat verantwoordelijk is voor het *doden* of vullen van het gat. De *levensverwachting* van het gat, oftewel *persistentie*, is dan gedefinieerd als het absolute verschil tussen waarden van dood en geboorte. Over het algemeen geldt dat hoge persistentie gaten gaten met lagere persistentie ingebed in zich kunnen dragen.

Persistentiehomologie wordt gepresenteerd met *persistentiediagrammen* (Edelsbrunner & Harer 2010), die uit een puntenwolk in $\mathbb{R}^2$ bestaan. Elk punt wordt geassoci-



Figuur 5 Een typisch persistentiediagram ontstaan uit de realisatie van een Gaussisch stochastisch veld. Dit is een puntenwolk waarbij ieder punt een uniek topologisch gat representeert. De coordinaten van het punt corresponderen met het kritieke punt verantwoordelijk voor de geboorte en dood van het gat.

eerd met een unieke topologische verandering in de variëteit.

Er is een diagram voor elke ambiente dimensie van de variëteit. De punten in het 0-dimensionale diagram leggen samensmeltingsvoorvallen van twee geïsoleerde objecten vast.

De punten in de 1-dimensionale diagrammen leggen de vorming en vernietiging van lussen vast; punten in 2-dimensionale diagrammen de geboorte en dood van topologische voids. Figuur 5 laat een typische persistentie zien van een massaverdelingsveld ontstaan uit een stochastisch proces.

Persistentiediagrammen bevatten een schat aan informatie. Het 0-dimensionale diagram legt samensmeltingen vast, die een karakteristieke eigenschap zijn van hiërarchisch evoluerende massaverdelingen.

De 1-dimensionale diagrammen coderen de topologische signalen van de percolatie-eigenschappen. De 2-dimensionale diagrammen coderen de hiërarchische aspecten van de vorming en samensmelting van voids. Bovendien wordt deze informatie op hiërarchische wijze gepresenteerd, zodat voids op kleine schaal met gemiddeld hogere dichtheid een bestanddeel zijn van grotere voids met lagere gemiddelde dichtheid. Dit heeft betrekking tot de recente ontdekking dat kosmische voids continu evolueren door samensmelting en andere complexe processen (Sheth & van de Weygaert 2004). Echter, een beoordeling van het exacte verband tussen topologische voids en kosmische voids is noodzakelijk.

5 Kosmische homologie en persistentie: resultaten

Dit proefschrift richtte zich met name op de topologische analyse van enkele kosmisch-relevante massaverdelingen door middel van homologie, Morse-theorie en persistentie. Hieronder geven we een samenvatting van de resultaten.

5.1 Convergentie van persistentiediagrammen en stochastische processen

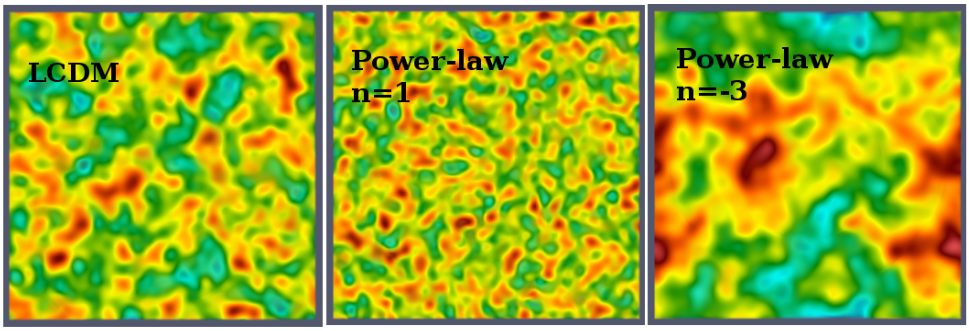
In Hoofdstuk 2 en Hoofdstuk 3 presenteren we empirisch bewijs dat het ensemble-gemiddelde van persistentiediagrammen ontstaan uit stochastische processen convergeert naar een stabiel gemiddelde. We laten dit zien door middel van een aantal representatieve voorbeelden van stochastische processen, zoals het Poissonproces, Gaussische stochastische velden, Voronoimodellen en het Soneira-Peebless model (later in dit hoofdstuk beschreven).

5.2 Homologie and persistentie van Gaussische velden

Gaussische velden zijn belangrijk voor de beschrijving van primordiale schommelingen in het dichtheidsveld. Een Gaussisch veld is geheel vastgelegd door het power-spectrum. In dit proefschrift analyseren we modellen gespecificeerd met zowel een machtswet power-spectrum alswel een LCDM spectrum.

Figuur 6 laat 2D plakken van een verwezelijking van een 3D Gaussisch veld zien, vastgelegd door verschillende power spectra. Naarmate de index van het power-spectrum afneemt worden de structuren zichtbaar groter.

In Hoofdstuk 3 analyseren we de persistentie topologie van de modellen. We laten zien dat het ensemble-gemiddelde van de persistentiekaarten van Gaussische velden stabiel zijn en goed-gedefinieerd. Als voorbeeld laten we in Figuur 7 het



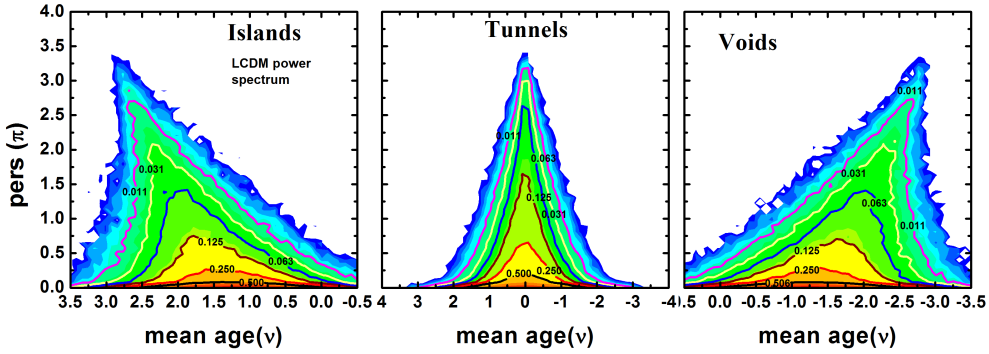
Figuur 6 2D plakken van een een verwezelijking van 3D gaussische stochastische veldmodellen. De modellen zijn geconstrueerd in een simulatiekubus met grootte $128h^{-1}$Mpc per zijde en grid-resolutie van $1h^{-1}$Mpc. Vervolgens zijn deze uitgesmeerd met een Gaussische kern met schaal $R_f = 2h^{-1}$Mpc. Het linkerpaneel laat een realisatie zien van het LCDM power-spectrum. De midden- en rechterpanelen laten realisaties van power-spectra gebaseerd op machtswetfuncties met indices $n = 1$ en $n = -3$ zien.

ensemble-gemiddelde van 100 verwezelijkingen van persistentiediagrammen van eilanden, tunnels en voids voor een LCDM spectrum zien. We tonen ook aan dat de kenmerken van persistentiediagrammen een directie weerspiegeling zijn van de singulariteitsstructuur van het veld. De karakteristieken van de corresponderende verschilkaarten hangen sterk af van de modelkeuze. Dit zou kunnen impliceren dat zij belangrijk kunnen zijn voor het onderscheiden van verschillende modellen. Dit is handig wanneer traditionele methodes hiertoe niet in staat blijken te zijn.

Bruikbare kwantitatieve informatie kan uit de marginale en cumulatieve distributie van de gemiddelde dichtheid en persistentie van de topologische kenmerken gehaald worden. We stellen vast dat de distributiefuncties een karakteristieke afhankelijkheid van de index van het power-spectrum vertonen. De marginale distributie en de gemiddelde dichtheid volgen een bijna-Gaussische verdeling. De marginale distributie van de persistentie impliceert een Poissonverdeling.

We duiken ook in de analyse van de Bettigetallen van Gaussische stochastische velden in Hoofdstuk 3. We vergelijken in dit kader de Bettigetallen met de vertrouwde Eulerkarakteristiek. We vinden dat de Bettigetallen een duidelijke afhankelijkheid laten zien van de index van het power-spectrum, dit in tegenstelling tot de Eulerkarakteristiekkromme, die dezelfde vorm heeft ongeacht de spectrale index. Een visuele indruk hiervan vindt men in Figuur 8, waarin we herschaalde Bettigetallen en Eulerkarakteristiekkrommes voor Gaussische stochastische velden presenteren.

Het verschil tussen de kenmerken van de Eulerkarakteristiek en de Bettigetallen heeft belangrijke gevolgen voor de beschrijving van de topologische structuur van dichtheidsvelden. Omdat de vorm van de Eulerkarakteristiek hetzelfde is ongeacht het power-spectrum zou men naïef kunnen aannemen dat de topologie binnen allle modellen hetzelfde is. Aan de andere kant overlappen de Bettigetallen over een aanzienlijk deel van de dichtheidsdrempelwaardes. De grootte van deze overlapping neemt toe met afnemende spectrale index. Dit wijst erop dat de topologi-



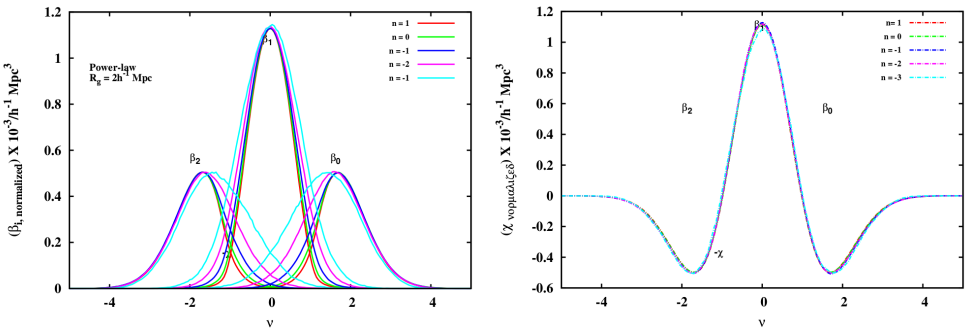
Figuur 7 Intensiteitskaarten van eilanden, tunnels and voids (van links naar rechts) in Gaussische stochastische velden voor het LCDM power-spectrum. Intensiteit I is gedefinieerd als de fractie van totaal aantal objecten (voor elke afzonderlijke dimensie) in een roostercel. Contourlijnen nemen toe in intensiteit met een factor 2.

sche structuur van het dichtheidsveld een mix is van verschillende topologiën voor een gegeven drempelwaarde. De hoeveelheid menging neemt toe met afnemende spectrale index, en laat zo de inherent verschillende topologische structuur van de verschillende modellen zien. Deze informatie wordt niet kenbaar met de Eulerkarakteristiek,

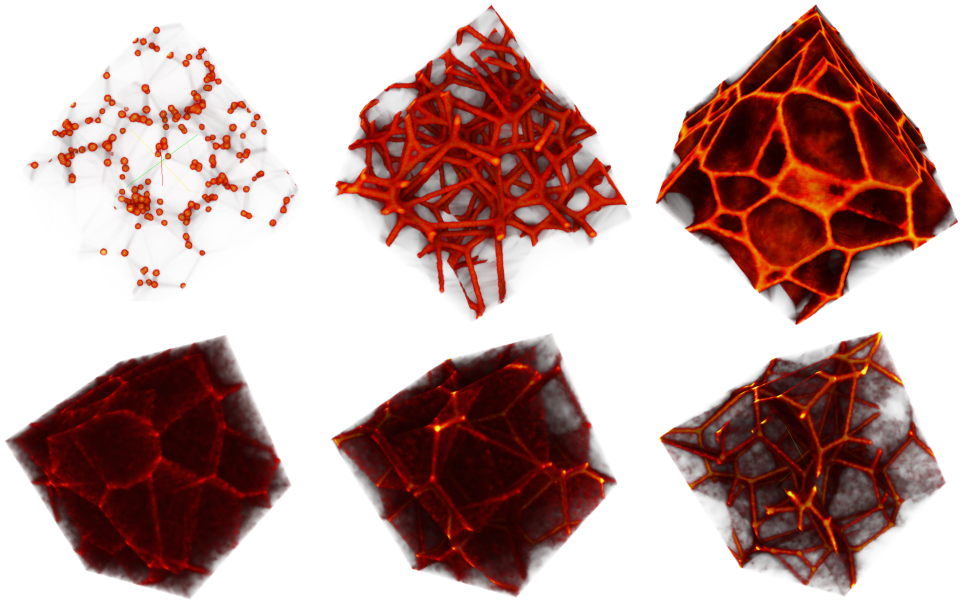
5.3 Homologie and persistentie van cellulaire verdelingen

Om de webachtige patronen van de groteschaalstructuur in het heelal te modelleren, vallen wij terug op *Voronoimodellen* (van de Weygaert 1991). Hierdoor proberen wij indicaties voor de verscheidene morfologische kenmerken van het web te identificeren. De modellen gebruiken Voronoitessellaties als matrijs voor de verdeling van massa en de hieraan gerelateerde melkwegstelselpopulatie. De *Voronoi-elementenmodellen* worden gekarakterizeerd door een dominante aanwezigheid van een enkel morfologisch element. De *Voronoi-evolutiemodellen* bootsten de evolutie van het heelal, op schalen van Megaparsec, na tot een webachtig patroon. De modellen die we hier gebruiken zijn behoorlijk geavanceerd, en representeren een vrij realistische uitbeelding van het kosmisch web voor kosmologieën gedomineerd door voids. De boven- en onderpanelen van Figuur 9 illustreren de rendering van het massadichtheidsveld voor een van de verwezelijkingen van de Voronoi elementen- en Voronoi-evolutiemodellen.

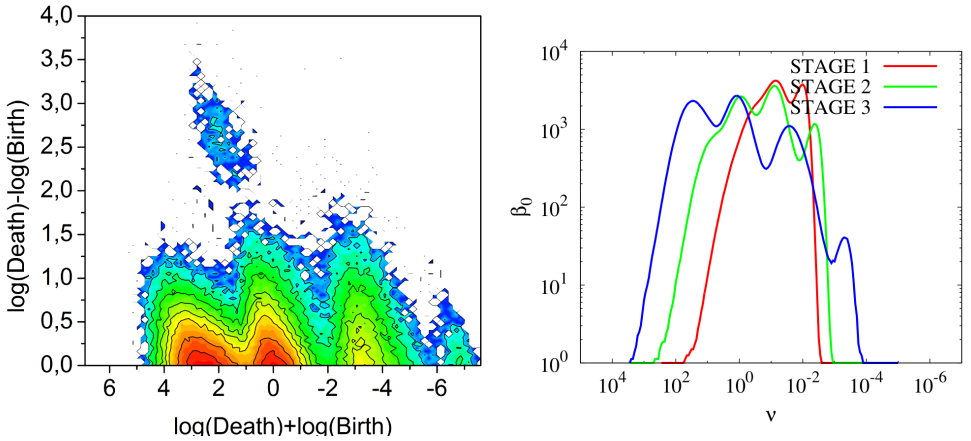
Voor de één-componentmodellen laten we zien dat de dominante aanwezigheid van verschillende morfologieën gereflecteerd wordt door kenmerken in de intensiteitskaarten in verschillende dimensies. Modellen waarin de deeltjes uitsluitend in en rond de knooppunten van het Voronoiskelet gedistribueerd worden laten de aanwezigheid van veel geïsoleerde clusters zien. Dit wordt goed vastgelegd door de geïsoleerde hoge persistentie wolken in de 0-dimensionale intensiteitskaarten. Modellen waarin de deeltjes uitsluitend rond de randen van het Voronoiskelet gedistribueerd worden, bootsen de filamentaire structuur van het web na. Onderling verbonden filamentaire patronen laten de aanwezigheid van goedgevormde lussen



Figuur 8 De krommes van herschaalde Bettigetallen (links) en de Eulerkarakteristiek (rechts) voor Gaussische stochastische velden, vastgelegd door een machtswet power-spectrum. De getekende krommes zijn voor spectrale indice $n = 1, 0, -1, -2 \& -3$. De krommes zijn herschaald met betrekking to de amplitude van de kromme voor het $n = 0$ model. De Bettigetallen laten een karakteristieke afhankelijkheid van het power-spectrum zien, in tegenstelling tot de kromme van de Eulerkarakteristiek. Dit betekent dat de Bettigetallen gevoeliger zijn voor de index van de machtswet dan de Eulerkarakteristiek.



Figuur 9 Dichtheidsveld corresponderend met de Voronoi-elementenmodellen (kader boven) en kinematische Voronoimodellen (kader onder). Van links naar rechts bestaan de elementenmodellen voornamelijk uit clusters, filamenten en wanden. De kinematische modellen evolueren van een fase waarin de meeste massa in de buitengebieden is tot een fase waarin de meeste massa zich in clusters en filamenten bevindt.



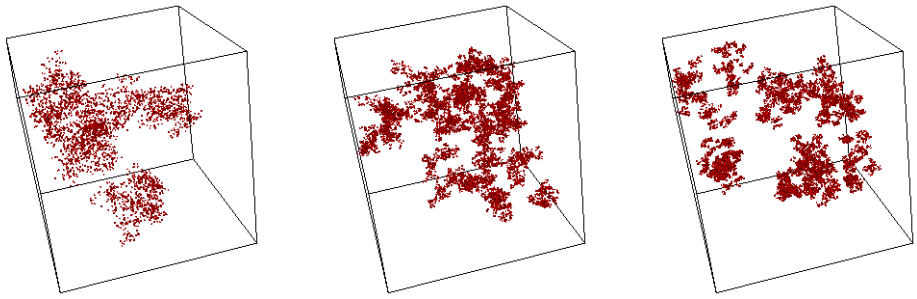
Figuur 10 Links: 0-dimensionale intensiteitskaarten voor de sterk geëvolueerde fase van een kinematisch Voronoimodel. Rechts: Bettigetallen afhankelijk van de dichtheidsdrempelwaarde voor drie verschillende evolutiefases van de kinematische Voronoimodellen.

zien. Dit wordt gereflecteerd in een hoge persistentie wolk in de 1-dimensionale intensiteitskaarten. Modellen waarin de deeltjes uitsluitend in de vlakken van het Voronoiskelet worden gedistribueerd bootsen de wandachtige structuur van het web na. Deze modellen laten de aanwezigheid zien van volledig omsloten voids. Dit aspect wordt vastgelegd door geïsoleerde hoge persistentie wolken in de 2-dimensionale intensiteitskaarten. We laten ook zien dat de dominante aanwezigheid van een enkele morfologie gereflecteerd wordt in de aanwezigheid van een enkele goedgedefinieerde mode in de grafieken van Bettigetallen in verschillende dimensies.

Voor de Voronoi evolutiemodellen laten we zien dat de topologische eigenschappen van de verschillende morfologieën in de verdeling – clusters, filamenten en wanden – goed gescheiden kunnen worden in de intensiteitskaarten. Bovendien wordt het aspect van evolutie vastgelegd door de evoluerende kenmerken in de intensiteitskaarten. Als voorbeeld laat Figuur 10 intensiteitskaarten zien voor de eilanden in de meest geëvolueerde realisaties van de kinematische Voronoimodellen. De kaart heeft vier duidelijke kenmerken, die op de aanwezigheid van de vier morfologieën duiden, namelijk, clusters, filamenten, wanden en voids. De aanwezigheid van meerdere en evoluerende morfologieën wordt ook gereflecteerd door meerdere goedgedefinieerde modes in de Bettinummerkrommes. Figuur 10 laat de Bettinummerkrommes zien voor drie verschillende fases van het Voronoi evolutiemodel (evolutie gaat van Fase 1 tot Fase 3). De aanwezigheid van meerdere modes, alsook een toename van het aantal modes voor meer-geëvolueerde modellen, is evident.

5.4 Homologie and persistentie van hiërarchicsche verdelingen

Om de kenmerken van meerschalige verdelingen te bestuderen, richten we onze aandacht op de Soneira-Peeblesverdeling (Soneira & Peebles 1978). Dit is een analytisch, self-similar, spatieel puntproces gedefinieerd met als doel het modelleren van de verdeling van melkwegstelsels. Het model behelst een hiërarchie van structuren



Figuur 11 Deeltjesverdeling in een bepaalde verwezelijking van het Soneira-Peeblesmodel. Het grootste deel van het volume is leeg door de hoge concentratie van deeltjes in een klein volume.

van verschillende dichtheden en karakteristieke groottes. Het wordt geconstrueerd door een aantal 'kinder'-bollen op willekeurige plaatsen in een 'ouder'-bol te plaatsen, die zelf weer op een willekeurige plaats in $\mathbb{R}^3$ geplaatst is.

De bollen op een hoger niveau corresponderen met dichtheidsstructuren op kleine schaal en de bollen op een lager niveau met lagedichtheidsstructuren op grotere schaal. Omdat iedere bol op dezelfde manier geconstrueerd is onstaat een self-similar en gebonden fractaal.

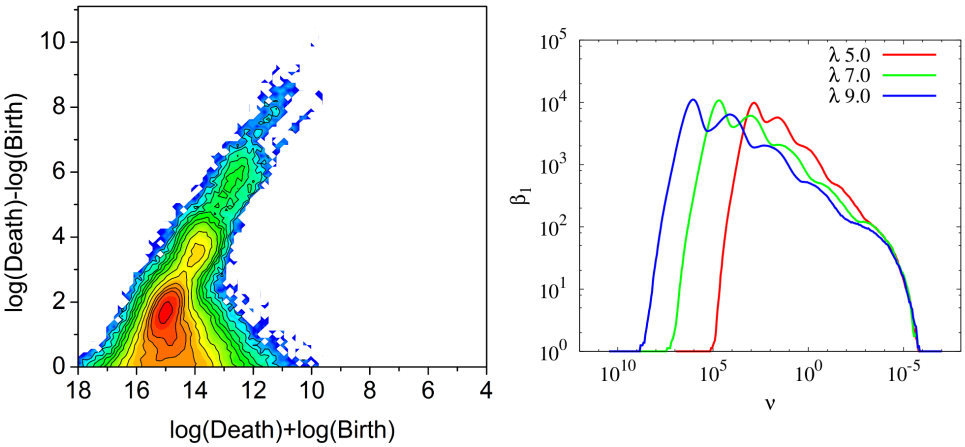
Een belangrijke eigenschap van dit model is, dat de statistische eigenschappen geheel en exact berekend kunnen worden. Hiermee bedoelen we de machtswet van de 2-puntscorrelatiefunctie, correlatiedimensie en de Hausdorffdimensie.

Figuur 11 laat een deeltjesverdeling zien in een bepaalde verwezelijking van het model, met drie verschillende concentraties van deeltjes. Het aantal deeltjes is gelijk voor iedere verwezelijking.

We laten zien dat de verschillende hiërarchische niveaus zich duidelijk openbaren in de intensiteitskaarten. In het rechterpaneel van Figuur 12 laten we een intensiteitskaart van een typisch Soneira-Peeblesmodel zien. De kaart laat de aanwezigheid zien van een aantal duidelijke klompen. Deze klompen zijn geplaatst op toenemende hoogte van hun persistentiewaarde. Het aantal klompen is gerelateerd aan het aantal niveaus in de hiërarchie van de verdeling.

De Bettigetallen leggen ook de verchillende hiërarchische niveaus in de verdeling vast. We zien dit in het rechterpaneel van Figuur 12. De meerdere toppen in de grafieken leggen het aantal niveaus in de hiërarchie vast. Voor hoge concentraties zijn de toppen duidelijk gescheiden, hetgeen betekent dat de niveaus in de hiërarchie ook duidelijk gescheiden zijn.

Voor lage concentraties zijn de niveaus niet zo duidelijk gescheiden, met als gevolg dat de deeltjes een diffuse verdeling hebben. Dit wordt gereflecteerd in het relatief lage aantal verschillende toppen in de Bettinummerkrommes.



Figuur 12 Links: Intensiteitskaarten van het Soneira-Peeblesmodel voor een hoge concentratie deeltjes. Rechts: Bettigetallen van het Soneira-Peeblesmodel for modellen met verschillende concentraties.

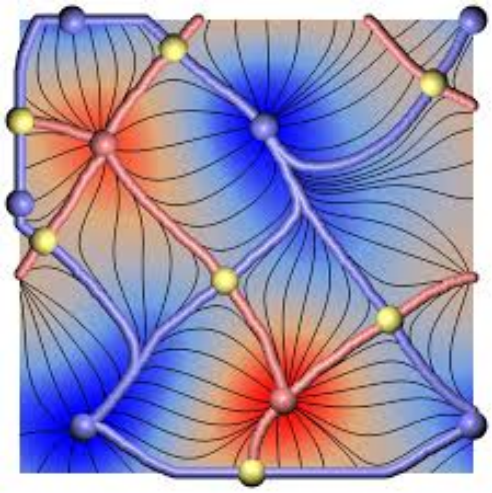
6 Op topologie gebaseerde visualisatie van kosmische filamenten: Felix

Het laatste deel van dit proefschrift introduceert en beschrijft een techniek voor de identificatie van filamenten, gebaseerd op de topologische eigenschappen van het dichtheidsveld. Het verkennen van de filamentaire patronen van het kosmisch web is een uitdaging vanwege het dynamisch bereik van de spatiële schalen en dichtheidsschalen ervan. Een goede karakterisatie moet ook rekening houden met de hiërarchische natuur van de structuren, iets wat behoorlijk moeilijk is.

Het verschil tussen ruis en betekenisvolle structuren is vaak slecht omschreven, en op verschillende punten kan ruis verward worden met daadwerkelijke structuren in de hiërarchisch evoluerende massaverdeling. Dit probleem is nog meer uitgesproken wanneer men de eigenschappen van ijle filamenten en wanden in void-achtige gebieden met lage dichtheid bekijkt. Een cruciale vereiste voor het begrijpen van hoe de vorming en evolutie van melkwegstelsels afhangt van het lokale milieu is door deze eigenschappen accuraat te classificeren. Deze eigenschappen omvatten ondermeer de dichtheidseigenschappen als de morfologische classificatie.

De techniek is gebaseerd op Morse-theorie en verloopt door het segmenteren van een variëteit gebaseerd op de karakteristieken van kritieke punten en de eigenschappen van de gradiënt van het scalaire veld dat hierop gedefinieerd is. Een belangrijk ingrediënt voor het begrijpen hoe de variëteit gepartitioneerd kan worden is het concept *integraallijnen*. Een integraallijn is een pad waarover de raaklijn van het pad parallel ligt aan de gradiënt ∇f op elk punt van het pad. Als voorbeeld presenteren we integraallijnen van een 2D scalaire functie als dunne grijze lijnen in Figuur 13.

De integraallijnen definiëren de oplopende en aflopende variëteiten van een kritiek punt. De verzameling van alle integraallijnen uit een kritiek punt p worden samen met p de oplopenden variëteit van p genoemd. Evenzo wordt de verzameling van integraallijnen die eindigen in het kritieke punt p met p samen een *aflopende*



Figuur 13 Morse-Smale complex van een 2D scalaire functie. De kritieke punten zijn aangegeven met bolletjes. De maxima zijn rose, zadelpunten geel en minima blauw. De dunne grijze lijnen die uit de kritieke punten komen en daar ook eindigen zijn de integraallijnen van de functie. De dikke grijze buisvormige structuren geven de separatrices aan, die de rand tussen twee aangrenzende klimmende en dalende variëteiten aangeven.

variëteit van p genoemd. De oplopende variëteiten (evenals de aflopende variëteiten) van alle kritieke punten segmenteren het domein.

Het *Morse-Smale complex* wordt geconstrueerd door de decompositie van de variëteit in meerdere cellen door de verzameling van integraallijnen die een gemeenschappelijke bron en bestemming hebben (Forman 2002; Edelsbrunner et al. 2001).

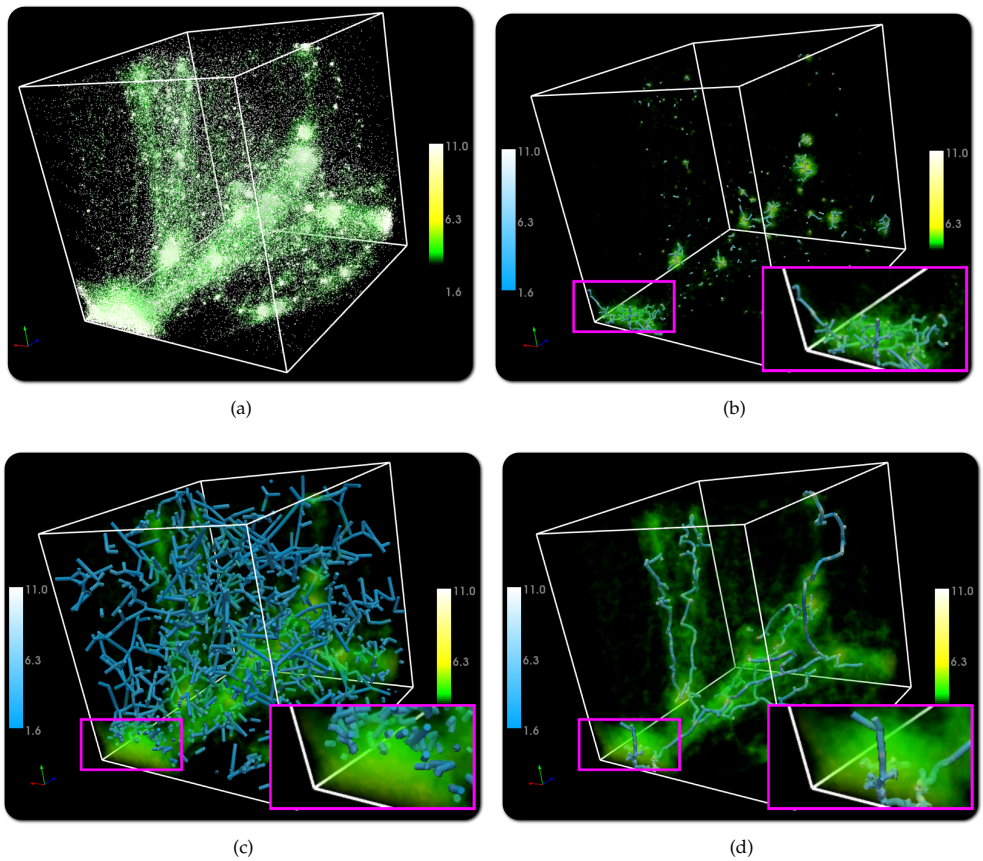
Dit werkt alleen wanneer de oplopende en aflopende variëteiten van alle kritieke puntparen elkaar slechts transversaal snijden. Elke functie f die hieraan voldoet wordt een *Morse-Smale functie* genoemd.

Figuur 13 laat een Morse-Smale complex van een 2D scalaire functie zien. De maxima zijn aangegeven met rose, de zadelpunten met geel en de minima in blauw. De buisvormige elementen representeren de bogen van het Morse-Smale complex die het domein segmenteren.

De verschillende elementen van het kosmisch web hebben een voor de hand liggende associatie met de elementen van het voor kosmische dichtheidsvelden geconstrueerde Morse-Smale complex.

Clusters kunnen worden geïdentificeerd met lokale maxima van het veld. Dit komt doordat massa uit alle richtingen naar de clusters toe stroomt. Dit komt overeen met het feit dat de gradient van het veld sterker wordt in de richting van het lokale maximum in het Morse-Smale complex.

De filamenten kunnen worden geïdentificeerd als de oplopende variëteiten van 2-zadelpunten. De oplopende variëteit van een 2-zadelpunt verbindt twee nabije maxima met elkaar en is een 1-dimensionaal element. De gradiënt langs de oplopende variëteit van een 2-zadelpunt neemt toe in de richting van de maxima. Dit impliceert een natuurlijk verband met kosmische filamenten, die ook quasi-lineaire structuren vormen en functioneren als de kanalen waardoor massa naar de clusters stroomt. Voids kunnen worden geïdentificeerd als de oplopende variëteiten van minima, waauit alle massa wegstroomt. Wanden kunnen geïdentificeerd worden als de



Figuur 14 Drie categoriën van filamenten uit de Cosmogrid dataset. (a) Deeltjesverdeling met een volumerendering van de DTFE dichtheid. (b) Filamenten in clusterachtige gebieden. De gemarkeerde gebieden laten het behoud van ingewikkelde topologische structuren binnen een groot clustergebied zien. (c) Filamenten in void-achtige gebieden, geëxtraheerd met parameters. De bijfiguur laat een clusterachtig gebied waaruit de filamenten weggefiltreerd zijn zien. (d) Filamenten die zich uitstrekken uit een clusterachtig gebied tot zelfs void-achtig gebieden.

aflopende variëteiten van 1-zadels die de voids scheiden. Deze zijn 2-dimensionaal van karakter.

We beschrijven hier Felix[12], een op topologie gebaseerd visueel interactief framework om filamentaire structuur uit een hiërarchie van Morse-Smale complexen van het dichtheidsveld te halen.

De filamenten in Felix worden geparametrizeerd door de dichtheidswaarden van de maxima en 2-zadels waardoor ze gedefinieerd worden. We maken een door de gebruiker gedefinieerd bereik mogelijk, zodat er meer flexibiliteit en accuratesse is bij het detecteren van de noodzakelijke eigenschappen. Dit in tegenstelling tot de op persistentie gebaseerde strategie geïmplementeerd in DisPerSe (Sousbie 2011),

waarbij lage-dichtheidseigenschappen onvermijdelijk weggefilterd worden.

Als voorbeeld bestuderen we het karakter van filamenten in drie verschillende dichtheidsregimes in LCDM simulaties. Het eerste regime betreft filamenten in hoge-dichtheidsgebieden rond compacte dichte clusters, waarvan we weten dat zij verantwoordelijk zijn als de transportkanalen van massa naar de clusters. Het tweede regime betreft de ijle lage-dichtheid filamenten gevonden in void-achtige gebieden. Het derde regime betreft filamenten die zich uitstrekken van een clusterachtig gebied naar een void-achtig gebied. Figuur 14 laat een voorbeeld van deze drie categorieën van filamenten, geëxtraheerd uit kosmologische simulaties, zien.

Gebruikmakend van Felix classificeren we in testmodellen melkwegstelsels als cluster/filament of niet. Van deze melkwegstelsels is de classificatie a priori bekend. Door het vergelijken van de classficatie door Felix met de a priori classificatie kunnen we laten zien dat we met hoge efficiëntie de originele classificatie kunnen terugvinden. We vergelijken Felix ook met enkele andere geometrische en topologische methodes, namelijk DisPerSE, MMF/Nexus en SpineWeb (Aragón-Calvo et al. 2010). Hierdoor laten we zien dat onze methode meestal even goed of beter presteert dan de andere methodes. Door gebruik te maken van Voronoi testmodellen, laten we ook zien dat een op persistentie gebaseerde simplificatie van het Morse-Smale complex, zoals geïmplementeerd in DisPerSE, niet volstaat om de gewenste filamenten te extraheren.

7 Conclusies

Het gebruik van topologische data-analyse van structurele patronen, in het bijzonder gebaseerd op persistentie, wordt steeds populairder in verschillende disciplines zoals beeldvormend medisch onderzoek, cartografie, landbouw, etc. Dit komt vooral door de gelijkenis in de eigenschappen van de problemen. Voor al deze disciplines worden de datasets steeds groter en ingewikkelder, of het nu om medische dossiers, genetische sequentiëring, neurale netwerken in het brein, astrofisica, geschiedkundige archieven of sociale netwerken gaat. Topologie is een ideaal gereedschap voor het analyseren van patronen in deze grote hoogdimensionale datasets. Dat komt omdat topologie vermoedelijk het enige formalisme is dat nauwgezet vormen en patronen kan definiëren in elke dimensie.

Dit proefschrift vormt een culminatie van een interdisciplinaire samenwerking tussen de gebieden kosmologie, wiskunde en informatika. Aan de kant van de kosmologie was het doel om de topologische beschrijving van de kosmische massaverdeling uit te breiden met Morse-theorie, homologie en persistentie. Aan de wiskundige kant heeft het geprobeerd de algemene eigenschappen en het karakter van persistentie, homologie en persistentiediagrammen te verkennen.

7.1 Toekomstideeën

Aan het einde van dit proefschrift, verschaffen we een aantal mogelijke richtingen om ons onderzoek op te volgen. De volgende niet-limitatieve lijst bevat mogelijke, veelbelovende onderzoeken. A

[12] De naam Felix is een afkorting van *Fil*ament *ex*plorer.

Statistiek van persistentie. We hebben empirisch vastgesteld dat het ensemblegemiddelde van een persistentiediagram van een stochastisch proces welgedefinieerd is. Een analytische probabilistische en statistische beschrijving is een vanzelfsprekende en belangrijke extensie van dit werk.

Persistentie (Homologie) van het grootschalig heelal. In dit proefschrift presenteren wij een topologische kenschetsing van Gaussische stochastische velden, met als doel het karakter van schommelingen in de dichtheid van het vroege heelal te begrijpen. Vervolgens analyseren we Voronoimodellen die de aspecten van de massaverdeling van het grootschalig heelal nabootsen. Een logische en belangrijke uitbreiding hiervan is het onderzoeken van de topologie van authentieke verdelingen van donkere materie en halo's in LCDM simulaties.

Persistentiekarakterisatie van kosmische achtergrondstralinganisotropie. Een op persistentie gebaseerde hiërarchische karakterisatie van de anisotropiën in de kosmische achtergrondstraling (CMB), zoals verstrekt door de laatste metingen van de PLANCK satelliet, biedt een interessant uitdaging. Persistentie heeft de mogelijkheid om de hiërarchische natuur van primordiale schommelingen aan het licht te brengen. Het zou ook interessant kunnen zijn om de CMB te onderzoeken met het doel bewijs te vinden voor primordiale niet-Gaussianiteit. Met bestaande methodes heeft het PLANCK team geen bewijs gevonden voor niet-Gaussische signalen in CMB kaarten. De hiërarchische beschrijving van persistentie heeft de mogelijkheid om dieper door te dringen in deze materie.

Filamentcatalogi van het heelal op grote schaal. Nadat wij de robuustheid van de filament-identificerende software hebben getest op testmodellen en simulaties, is het belangrijk om een catalogus van filamenten op te bouwen, zowel uit simulaties als waarnemingen. Deze catalogi bieden de mogelijkheid om de eigenschappen van de geassocieerde melkwegstelses op te vragen, om de vorming en evolutie van melkwegstelsels te bestuderen vis-a-vis de grote schaalstructuur waarin zij zich bevinden.

Hiërarchische kenschetsing van kosmische voids. 2-dimensionale persistentiediagrammen die de vorming en evolutie van topologische voids vastleggen, kunnen mogelijk een belangrijk gereedschap zijn voor de kenschetsing van de hiërarchische aspecten van kosmische voids. Dit is in het bijzonder interessant in het licht van het feit dat topologie miscchien wel de enige methode is die een void uniek en ondubbelzinnig kan definiëren zonder additionele vrije parameters. Sinds lange tijd werd theoretisch vermoed hetgeen recentelijk observationeel is bevestigd: kosmische voids vormen en evolueren hiërarchisch. Verschillende studies laten zien dat voids gevoelig zijn voor donkere energie en mogelijk informatie bevatten over alternatieve zwaartekrachtstheoriën.

Kenschetsing van het reionisatietijdperk. Door het hiërarchische karakter van persistentie is het op natuurlijke wijze toegesneden voor de studie van reionisatie en het evoluerende netwerk van ionisatiebellen in het ionisatietijdperk, dat het begin van

de vorming van sterren en melkwegstelsels markeert. 0-dimensionale persistentiediagrammen zijn optimaal toegelegd voor het vastleggen van het evoluerende netwerk van ionisatiebellen, die in essentie samensmeltende geïsoleerde objecten zijn. Bovendien zal het bestuderen van de eigenschappen van 1-dimensionale diagrammen veel details blootleggen over de percolatie eigenschappen van de verdeling in dat tijdperk. De 2-dimensionale diagrammen bieden de mogelijkheid om het evoluerende hiërarchische netwerk van voids op jonge leeftijd te onthullen.

Acknowledgements

I arrived in Groningen on a cold, dreary winter day afew years back. Many winters later, and quite a few experiences in my bag – some extraordinarily beautiful, and some not so – I am at the finishing line of the primary goal of this journey. My journey in to science started on quite a random note, nut I can say with conviction now that this random foray has been one of the most amazing things to have happened to me. The success of this journey is by no means a doing of my own exclusively. If not for the unwavering belief that numerous people showed in my ability, and the unflinching support that they have given me, this gargantuan task would have been like a flower nipped in the bud. I take this opportunity to thank everyone who has been a part of this saga – visibly or invisibly.

For reasons best known only to me, the question whether to start with my parents or my brother, has been running in the background in my mind for almost a year now. The resolution has come only towards the end, when I spent the last few months with my parents and rediscovered them.

Ma and papa, words fail me when I attempt this, cheifly because I am not adept at things emotional and such. Despite my handicap, I am in a constant awe and reverence by how you have chosen to live your life. Your unflinching devotion towards living a moral life, and doing the right thing, even in the face of the strongest adversities is a source of highest inspiration. I know I cannot do much wrong, even if I am tempted to sometimes, because I have you as my example. You give me hope that the Universe looks after its good children. Your unconditional love is the stong pillar I hold on to, the solid rock on which I stand. If there is an iota of truth about the theories of reincarnation, do accept me as your child in the coming lives too!

Pushkar, biologically you are my elder brother, but technically nothing short of a second father. You filled my childhood with untold joy and imagination. I still remember how I used to look forward to returning from my boarding school during the vacations with an almost insane glee. This was all in an anticipation of hearing about your latest exploits in the magical kingdom of which you were the king. Not

to mention that funny and imaginary ghost of a slave you had, and your promise of making him my slave for a few rupees. It is another story altogether that despite having parted with all my childhood savings, and a constant imploring for the last 20 years now, I am yet to get a glimpse of the legendary *paad-foosooki*. On a more serious note, you have been my lord protector, and I always knew you were standing behind me whenever things got tricky. Always be the boss you are. I could not have asked for a different elder brother.

Rien, I have slowly come to realize that going through a PhD is equivalent to going through the motions of a relationship. There are many shades to it, and not all bright always. In the last couple of years, I had many "healthy debates" with you. But let there be no doubt that I have immensely valued your opinions and contributions, not only for the PhD, but also for life. If at times I have taken the liberty of expressing myself, sometimes vociferously and aggressively, it was only because somewhere I knew you have all good things for me in your heart, and that you would take them in the right spirit. People tend to take liberties with people they are more comfortable with. This thesis would not have been anywhere close to what it is right now, if not for your constant nudges in the right direction.

Herbert, you have been a friend, philosopher and guide in every sense of the phrase. To top it up, you have been a saviour at a number of occasions – more than I could possibly keep a count of. My topological aspirations would have died a premature death, in the absense of that fateful encounter with you. I remember that conference in Avignon, France, when I met you and told you all topology passes miles above my head, mostly because the vocabulary is alien to me. Your characteristically demure reaction was to invite me for lunch. Over that lunch, and a few bottles of wine, you asked me to erase all I had learnt from my mind, and started drawing things in air. In those couple of hours, I learnt from you what I could not learn in six months in a class. I have not looked back since then. Thank you, Sir, for being there whenever I needed it the most.

Gert, to be honest, I was scared of you right from the beginning. It did not help things that my earliest encounters with you in the topology lectures, in the early sleepy winter mornings. It is my loss that I let that intertia, driven by fear, to avoid crossing crossing paths with you, unless absolutely necessary. However, I believe I have the years in the future to make up for it. I am aware you have been watching my progress silently and almost invisbly. I hope I give you a reason to be proud of me.

Bernard, it is only towards the end that I start to understand your role in my PhD. You have been there with an encouraging word constantly. A conversation with you can lift up the spirits of the most morose man. You mince no words in putting forward your opinion, yet you care deeply in your heart for your students. you are an immensely wise soul, with an infectious energy. It is not everybody's cup of tea to simultaneously juggle the various things that you do. I hope to keep picking your mind for many years to come!

Chanda Jog and Tarun Deep Saini deserve a special mention here, and for similar reasons. Chanda for being instrumental in pushing me towards astronomy, and reposing a faith in me against all advice. Thank you Mam, I have not met a more

kind soul yet. You are one of the finest scientists I know. Tarun, by your own admission, you advocated for me when everybody else was grappling with the idea if it was a risk. I hope I have stood up to that faith. You are more a friend, than a teacher.

A big thanks also to the lovely people at the Kapteyn institute. In this serie, a big special thanks goes to my *ouwehoeren* partner at the happy hours, Joost Adema. You are a great soul, and an extremely tall man, in your thoughts and physique. Life at Kapteyn would not have been the same without you.

Chiara, what can I say "mother"? Moms do not need to hear much, except perhaps "I love you".

Mark, my brother from another mother, I have always wondered how two people sharing the same birthday can have so radically different personalities. But life is long, and so is our friendship hopefully. My dilemma will resolve eventually.

Katinka, you have been an all-weather friend. I cannot forget you offered me a shoulder to cry in my darkest hours. It was novel to me, because crying was novel to me.

Filippo, this is perhaps the beginning of a great friendship. Keep growing in life and otherwise, if you know what I mean.

Thank you, Edwin, for all the chats about life and such. Your zeal for the things you care is exemplary. I hope Infoversum reaches the great heights it is destined to.

Jan, Andrey and Stephen, thank you for keeping me company during the insane friday evenings. Would not have been the same without you. Jan, the endless debates and philosophies on the human nature were refreshing, made more rosy by the beer of course. Andrey, the prospects of a *vodka-duel* still cross my mind. We may just have to postpone it to another life – I am getting older, and you are still a Russian.

Lucia, Hennie and Christa, thank you for being a ray of hope in my ever chaotic life. Without you, I would never have the peace of mid or the faith that I do not need to worry about anything beyond the academics.

Suzanne and Neel, thank you for making my life bearable in Austria. You kept me from getting insane in the closed confines of the institute in the remote Vienna woods. If not for you, I would have turned into a derelict ghost, wandering insanely in the Vienna woods.

I cannot afford to miss to mention another mother, and a band of brothers. I was ten when I left home and my mother, to venture out in the world. The pain of leaving my mother did not last long, because I discovered another mother, my *Vidyapith* – my school, along with a bunch of equally lost kids. It is a couple of decades later, and still our bond grows stonger by the day. Param, PT, KG, Dilip, Ashish, Baba and all the sixty-odd guys, I am lucky to have you around. My own mother thinks sending me away was the worst mistake she made. She could not be more wrong.

Thanks also to the "underground" of Groningen, and my non-scientist local friends (who are scientists of life in their own way) deserve a special mention too. You never for once let me feel I was an outsider. It is to your credit that I had the most awesome time in Groningen and consider it my second home. Annie, Mit, Yorick, Roger, Joy, Rutger and Remco you deserve a a special mention here.

Meenal, I do not know you for long. But thank you for being a part of our family, and bringing joy to us all. You know what I am talking about. You also make my

brother happy, and that makes me happy.

Last but not the least, thank you for being in my life Maryse. God knows this thesis would literally have been impossible without your contributions. You have been there to support me through every rough patch. You showed me the beauty of a soft heart and a warm smile. I am glad I met you!

I also thank the Universe for just being there!

Pratyush Pranav
Delhi, November 2015